15
VOICE AND SPEECH FUNCTIONS (B310–B340)

Elspeth McCartney

What is the construct?

The International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY) domain 'voice and speech functions' (b3) includes production and quality of voice (b310), articulation functions (b320), fluency and rhythm of speech (b330) and alternative vocalizations (b340, such as making musical sounds and crying, which are not reviewed here). The underpinning construct is integrity or deviation of functional mechanisms of speech production. Overviews of ICF categories within this domain have been undertaken by Ma and colleagues[1] for voice impairments (without the CY update); by McLeod and McCormack[2] for child speech functions (including the CY update), and by Yaruss and Quesal[3] with respect to fluency (without the CY update). These authors list the many ICF codes relating to body structures (such as s240–260 structures of the ear and s3 structures involved in voice and speech), other body functions (such as b1560 auditory perception, b167 mental functions of language, b230 hearing functions and b440–445 respiratory system), activities and participation (particularly d3 communication), and environmental factors that may be associated with voice, speech and fluency impairments (and see also relevant chapters in this volume).

Those interested in the ICF-CY/b3 include speech–language pathologists (SLPs) offering clinical services to children diagnosed with voice or speech disorders or dysfluency. The ICF-CY does not consider the aetiology of health functions, which are classified by the International Classification of Diseases (ICD-10).[4] Voice or speech disorders or dysfluency may occur in isolation as a primary difficulty for a child, but clinicians are also interested body functions secondary to known aetiologies. For example, children with Down syndrome often show difficulties in voice, speech and fluency.[5,6] An emphasis on clinical categories has tended to slant outcome measurement towards assessments that differentiate children with impairments from typically developing children. For example, although it is possible to code fluency, rhythm, speed and melody of speech separately (b3300–b3303), and any of these may be relevant to an individual child, most research interest has centred on differentiating stuttering from normal developmental non-fluency, and in assessing reductions in stuttering after treatment. Fluency is therefore more likely to be assessed than speech melody. Features targeted in therapy are most frequently measured after research and clinical studies of treatment efficacy. Children are often assessed on entry to therapy and reassessed at the end of an episode of care using the same instruments, allowing 'before and after' comparisons.

Recent prevalence statistics from a survey of parents of nearly 5000 4- to 5-year-old children weighted to represent the Australian population reported that 2.2% of their children had voices that sounded 'unusual', 6.0% had speech 'not clear to the family', 12% had speech that was 'not clear to others' and 5.6% 'stuttered, stammered or lisped'.[7] Teachers of 10 425 children aged 4 to 12 years in 36 primary schools in Sydney, Australia, reported prevalence rates of 0.12% for voice disorders, 1.06 for speech disorders and 0.33% for stammering, using standardized descriptors and confirmed where possible by SLP reports.[8] Overall, prevalence estimates decreased with age. Children requiring investigation and intervention for voice, speech or fluency form a significant part of the caseload for some SLP services that deal with younger children. For example, a cohort of children mostly aged 2 to 4 years referred to a paediatric UK SLP service over a 16-month period showed 2.0% with voice or nasality disruption, 29.1% with speech difficulties and 5.3% with dysfluency.[9]

For young children, particularly those with primary voice, speech or fluency impairments, intervention is often directed towards improving function with either a curative (normalization) or a habilitative (improvement) care aim, anticipating communication benefits. Primary voice dysfunction may be associated with vocal misuse and improve with therapy. Speech disorders typically manifest and are treated in the pre- and early school years, with intervention aimed towards the production of adult-language speech sounds. Developmental stuttering similarly is evidenced and treated in early childhood, although it may persist into adulthood, and for young children intervention aims to reduce moments of stuttering and improve fluency.

Where impairments persist into later childhood, however, intervention may move towards enabling, that is optimizing the use of existing functions, and activity and participation measures become increasingly relevant. These are reviewed under 'Communication' in Chapter 22.

General factors to consider when measuring this domain

One ICF code – articulatory functions (b320) – with no subcategories has to serve for all disorders of speech function. It includes functions of enunciation and articulation of phonemes; spastic, ataxic and flaccid dysarthria; and anarthria. As no other ICF-CY codes are available, no distinction is made between phonological substitutions and speech affected by motor impairments: McLeod[10] suggests additional codes that could usefully be added. Although speech terminology is far from standard,[11] 'articulatory' as the title of the ICF code for all speech difficulties may be confusing to some SLP practitioners.

Clinical and Research Applications of Outcome Measures

Two aspects of 'child talk' affect the measurement of voice, speech and fluency: inconsistency in the occurrence of disruption, and the ephemeral nature of the speech signal. Phonological output patterns are usually consistent at a whole-word level for typically developing children,[12] and only around 10% of referred children with primary speech impairments are reported to show inconsistent speech,[13] so a brief speech sample will either evidence a child's consistent level of functioning and so form a valid basis for analysis, or show that further systematic sampling is needed [and see the review of the Diagnostic Evaluation of Articulation and Phonology (DEAP), below]. Vocal function may be inconsistent depending upon fatigue and voice-use factors, and so will require repeated sampling. The occurrence of stuttering is highly inconsistent across talk samples, showing fluent speech or overt stuttering even within a brief time interval.[14] The amount and type of dysfluency may be influenced by child, linguistic and contextual factors that are highly individual and may vary across a child's development. Obtaining a 'representative' sample of talk to assess fluency will therefore require careful thought, and several samples from a variety of contexts will be needed for analysis [and see a review of the Percentage Syllables Stuttered (PSS) below].

As talk is ephemeral, rapid transcription and data-gathering techniques are required, using either live or recorded speech samples. These have to be checked for reliability and subjected to detailed analyses. Data collection and analysis requires a 'trained ear' and technical skills, for example in evaluating voice attributes or using the International Phonetic Alphabet (IPA). Such skills are acquired during SLP training, involving a considerable investment in time and expertise. Once acquired, inexpensive data-gathering and analysis protocols can show reliable and valid intra- and inter-rater results (see reviews below). Nonetheless, demonstrating rater reliability continues to be problematic for some measures, such as identifying moments of dysfluency, characterizing voice disorders or capturing fine-grained phonetic detail. As noted in reviews, some measures have been shown to be reliable with trained users or within research studies, but, although used in general clinical practice, their reliability in that context has not as yet been firmly established.

As well as data-analysis protocols, standardized measures and commercially published assessments have been developed that should offer psychometrically adequate standards of reliability and validity. However, these are not always demonstrated. Widely used measures of child speech and fluency were analysed, amongst other instruments, to determine their psychometric quality by the US Agency for Healthcare Research and Quality (HSTAT 52):[15] voice measures not specifically adapted for children were also included. In general, child assessment measures fared badly. The reviewers noted that reliability and validity data rarely came from peer-reviewed literature, but rather from publishers' manuals accompanying assessments. It is therefore not safe to assume that all published measures will show acceptable psychometric characteristics.

Summary

Analysis of core voice and speech functions may be applied to naturalistic samples of child talk, and are thus applicable to a wide range of clinical contexts and are inexpensive to administer by trained SLPs. They do however have inbuilt reliability issues, associated with the ephemeral nature of speech and with sampling difficulties. Since reliability is a precondition of validity and the measurement of change, this presents a continuing challenge to the development of clinical outcome measures. Standardized measures of core functions should have overcome reliability and validity problems before publication, but require individual scrutiny to establish whether this is the case. In this chapter, two instruments are included for each of voice, speech and fluency functions – (1) one clinical data-collection and analysis protocol to assess core function, and (2) one commercially packaged measure. These have been selected as illustrative only, but are measures currently in wide clinical use.

Overview of recommended measures

Voice Functions (b310)

Consensus Auditory-Perceptual Evaluation of Voice

Overview and purpose

The Consensus Auditory-Perceptual Evaluation of Voice, fifth edition (CAPE-V),[16,17] was developed by the American Speech–Language Hearing Association as a tool for clinical auditory-perceptual assessment of voice. It measures severity based on a minimal set of voice parameters determined by a conference of international voice scientists and SLPs to be meaningful for clinical use.[17] It is designed to obtain reliable results expediently while offering a refined analysis of vocal function. SLPs define and rate six vocal attributes that are salient in identifying voice dysfunction: overall severity, roughness, breathiness, strain, pitch and loudness. Additional features may also be noted as relevant.

Administration and scoring

Three child voice samples from different tasks are audiotaped using standardized procedures and, if possible, entered into a computer. Task 1 records ‘vowels “a” and “i” ’ sustained for 3 to 5 seconds, each repeated three times. Task 2 involves the child reading (or younger children imitating) six standard ‘sentences’ to elicit (1) all vowel sounds, (2) easy voice onset using the letter ‘h’, (3) only voiced phonemes, (4) hard glottal attack, (5) many nasal sounds and (6) many voiceless plosives. Task 3 records at least 20 seconds of natural ‘speech’ offered in response to a standard question.

All three tasks are performed by the child before the SLP enters a score. Severity is scored for each vocal attribute on a 100-mm visual analogue line, where 0 equates to normal and 100 to severe dysphonia. Other points are scaled in between as ‘mildly deviant’, ‘moderately deviant’ and ‘severely deviant’. Severity judgements are made by the SLP and marked on the voice analogue line, and expressed both as a percentage and with the descriptive severity rating. If performance is uniform in severity across the three tasks, one severity rating is scored for each vocal attribute. If tasks show discrepant severity, each vocal attribute is recorded on the same scale but coded as Task 1, 2 or 3. Responses are also judged as either consistent, or intermittent if the voice attribute does not occur on each task but is of equal severity when it does occur. CAPE-V ratings may be repeated to assess treatment outcomes.

Simulations and practice, including child voice disorder examples, can be accessed at http://engage.doit.wisc.edu/sims_games/showcase/speechpathology/index.html.

Psychometric properties

Validation studies are ongoing, and current psychometric information comes mainly from adults. Braden et al[18] found a weak correlation between CAPE-V and an institution-specific self-perception scale of voice impairment only for moderate to severe dysphonia in a retrospective study of 199 adult clients who had received botox injections, and Eadie and Baylor[19] found significant correlations with acoustic measures.

Establishing rater reliability has been an ongoing problem in developing adequate auditory-perceptual scales to evaluate voice quality.[20] Rater reliability measures of CAPE-V using adult voice samples show effects of training and prompting. Sixteen inexperienced SLP [134]
graduate students judged sustained ‘a’ and recorded a standard speech passage by 54 speakers (48 dysphonic, six normal) using visual analogue scales, and were then trained using definitions, auditory ‘anchor’ examples of representative disorders and feedback on accuracy.[21] Inter-rater reliability improved with training. Karnell and colleagues[22] found intrarater Spearman’s correlations of 0.88 to 0.93 and inter-rater reliability of 0.86 to 0.93 on overall severity ratings by four experienced SLPs who [133]
had listened to ‘anchor’ samples before commencing rating sessions. Forty inexperienced SLP undergraduate judges[43] compared written definitions (‘written anchors’) with auditory anchors. After 20 minutes of training, judges rated 36 sustained vowels exemplifying normal, breathy, hoarse or rough voices when randomly allocated to a no-anchor, written-anchor, auditory-anchor or combined auditory–written-anchor listening condition. Provision of anchors significantly improved inter-rater reliability. Auditory anchors allowed more improvement than written, but combined written–auditory examples offered the strongest measure of improvement. These studies suggest that inter-rater reliability is strongly affected by training and provision of external standard comparison measures.

Three SLPs with experience of using CAPE-V investigated inter- and intrarater reliability of CAPE-V sentences with children,[44] rating audio-taped samples of 50

Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)	
Purpose	To provide an expedient clinical auditory-perceptual assessment of voice production and quality. To measure a minimal set of voice parameters and severity for clinical use
Population	Ages 4y to adult who are dysphonic or who are in the process of diagnosis
Description of domains (subscales)	One domain/six voice-quality subscales: severity, roughness, breathiness, strain, pitch, loudness. Additional features noted as relevant
Administration and test format	Time to complete: about 5min Testing format: child voice samples, audio-taped undertaking three tasks – sustained vowels, repeating standard sentences, natural speech Scoring: record each voice quality feature on 100-mm visual analogue line from three tasks combined. 0=normal–100=severe dysphonia. Also note any variation of attribute across tasks and any variation in severity Training: designed for qualified SLP use. No additional training
Psychometric properties	Normative sample: none *Reliability* Internal consistency: no item analysis; variation anticipated. Test–retest – no information retrieved; variation anticipated. Rater – shows effects of training and listening prompts (adult samples). Child samples – intrajudge intraclass correlation coefficient 0.62 (strain) –0.88 (breathiness), interjudge intraclass corellation coefficient 0.35 (strain), –0.71 (breathiness) *Validity* Content validity: expert consensus reported, with CAPE-V voice-quality features being those consistently assessed as 'clinically meaningful'. Construct/discriminant validity – correlations 0.89 (breathiness)–0.95 (severity) (adult voice samples) with Grade, Roughness, Breathiness, Asthenia, Strain: CAPE-V's visual analogue scale probably more sensitive to small differences. Significant correlations with acoustic measures Responsiveness: showed positive change following laryngeal reinnervation, adults
[144] How to order	Download from www.asha.org/NR/rdonlyres/C6E5F616–972F-445A-AA40–7936BB49FCE3/0/D3CAPEVprocedures.pdf or from Kempster et al (2009), below
Key references	ASHA Special Interest Division 3, Voice and Voice Disorders. *Consensus Auditory-Perceptual Evaluation of voice (CAPE-V)*. Available at: www.asha.org/NR/rdonlyres/C6E5F616–972F-445A-AA40–7936BB49FCE3/0/D3CAPEVprocedures.pdf (accessed 3 June 2009). Kempster GB, Gerratt BR, Verdolini AK, Barkmeier-Kraemer J, Hillman RE (2009) Consensus Auditory-Perceptual Evaluation of Voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol/Am Speech Lang Hearing Assoc* 18: 124–132.

children aged 4 to 20 years with airway conditions who had undergone at least one major laryngotracheal reconstructive surgery. No training was given, as the study aimed to generalize findings to the larger community of experienced voice-rater SLPs. Intrarater/inter-rater reliability interclass correlation coefficients were as follows: overall severity 0.86/0.67, roughness 0.86/0.68, breathiness 0.88/0.71, strain 0.62/0.35, pitch 0.86/0.68 and loudness 0.80/0.57. The low correlation for strain was considered to be due to the absence of visual evidence of excess vocal effort that would be available in live ratings of CAPE-V, and loudness may have been affected by listening conditions, despite standardization. The other four measures achieved moderate to strong correlations. On adult voice samples, comparison between CAPE-V and another widely used clinical measure of voice quality, Grade, Roughness, Breathiness, Asthenia, Strain,[25] showed high correlations – Spearman's coefficients 0.89

(breathiness), 0.95 (severity) – suggesting concurrent criterion-related validity.

Lee et al[26] showed CAPE-V scale improvements in severity, roughness and breathiness in 17 adult samples after reinnervation for unilateral vocal cord paralysis, suggesting responsivity.

Boone Voice Program for Children, second edition

Overview and purpose

The evaluation section of the widely used Boone Voice Program for Children[27] aims to provide a rapid voice screen to determine whether there is a problem, a detailed evaluation to plan therapy and a checklist to monitor improvements. The screen comprises a 'Voice Rating Scale' measuring eight voice parameters, and the 'S/Z ratio' that compares the length of time a child can sustain the unvoiced phoneme 's' and the voiced phoneme 'z' as an indication of laryngeal involvement.

If in-depth investigation is required after screening, the 'Voice Evaluation Form' assesses 10 aspects, of which four include voice. (Non-vocal aspects are not reviewed here.) The four voice-related functions measure nasal resonance, respiration, phonation and eight vocal parameters (breathing, pitch, pitch inflections, loudness, voice quality, horizontal focus, vertical focus and nasal resonance). If a therapy programme is undertaken, facilitating approaches are offered and changes in function are marked on a checklist to show outcomes.

Administration and scoring

The 'Voice Rating Scale' assesses a speech sample elicited by telling a child a story with supporting pictures and asking for a re-tell. Two stories are provided, one for children up to 9 and the other for children aged 10 to 12 years. The SLP decides, on listening, whether the child's voice sounds like that of their peers. The eight voice parameters are scored as negative (–), normal (N) or positive (+). Negative (–) scores represent too few words on a breath, low pitch, monotonous pitch, inadequate loudness, breathy voice, anterior-focused voice (infant voice), laryngeal-focused voice and insufficient nasal resonance. Positive (+) scores are given for too many words on a breath, high pitch, excessive pitch variation, excessive loudness, harsh/tight voice, excessive oropharyngeal resonance, nasal-focused voice and excessive nasal resonance. Otherwise, the score is normal. Any non-normal score suggests further voice evaluation.

The 'S/Z ratio' is assessed by demonstrating a prolonged 's' and then asking the child to sustain 's' for as long as possible, giving two trials. Prolongation time in seconds is recorded using a stopwatch, then the procedure is repeated using 'z'. The longer prolongation time for each consonant is used, dividing the longer 's' by the longer 'z', which is expressed as the S/Z ratio. An S/Z ratio >1.4 suggests laryngeal involvement.

Some sections of the 'Voice Evaluation Form' are supported by or require the use of instrumental measures as available in specialist clinical settings. Only the three non-instrumental assessments are outlined here – see Colton et al[28] for a review of instrumental measurement. 'Section 5 – Nasal Resonance' uses two stories read aloud by the child (or repeated by younger children). One contains no nasal consonants, and therefore a child perceived as nasal is generally hypernasal. The other has many nasal consonants and includes 25 all-oral consonant words. Denasality will be evident, and also assimilative nasality, where the 25 all-oral consonant words will sound nasal in this context, although not if read in isolation. 'Section 6 – Respiration' uses clinical observation to assess respiration using a prepared checklist, and the S/Z ratio may be repeated. 'Section 8 – Voice Rating Scale' involves a more finely graded analysis of the Voice Rating Scale than the screen, with the '–/N/+' judgement replaced by a seven-point scale.

Psychometric properties

No psychometric data are given, and norms for children are not given in the manual. However, measures are validated from earlier voice assessments (and see CAPE-V above). Good test–retest and inter-rater reliability for S/Z ratios have been shown in typically developing 6- and 7-year-old children.[29] Ninety-five per cent of adults and children combined with vocal fold margin pathology (nodules and polyps) showed S/Z ratios >1.4, whereas normal and dysphonic speakers without pathology approximated 1.[30] This justifies the Boone Voice Program cut-off. However, for 123 dysphonic children aged 5 to 15 years, 69 with vocal cord nodules and 54 without, S/Z ratio performance did not discriminate those with fold margin pathology,[31] perhaps owing to typical differences in nodule characteristics and size between adults and children. This suggests that the S/Z ratio is not a safe indicator of vocal fold margin pathology in children.

Articulation Functions – Phonology (b320)

Percentage Consonants Correct – revised

Overview and purpose

Consonants are usually differentially affected in comparison to vowels in functional speech disorders. The 'percentage of consonants produced correctly' [Percentage Consonants Correct measure (PCC[32])] in a speech sample measures the severity of speech dysfunction and may be used repeatedly to track speech outcomes. The PCC acts as a major predictor of severity ratings made independently by SLPs[33] and can be calculated from brief talk samples by a qualified SLP.

However, the original version of PCC devised by Shriberg and colleagues[33] marked subphonemic distinctions (consonant distortions) as errors. Such detailed analysis presents inherent difficulties for interjudge reliability, and so limits clinical usefulness. The measure was therefore revised by defining clinically common and uncommon consonant distortions[34] and counting both types as correct realizations. The revised version (PCC-R)[34] showed increased inter-rater transcription reliability, owing to the broader level of analysis. Sensitivity to true speech impairment was also increased by ignoring subphonemic distortions, which are frequent in the speech of young children, but retaining phonemic substitutions and omissions that are salient markers of speech delay and disorder.

Administration and scoring

Administration has not changed from the original paper,[33] where the procedures are detailed, although scoring procedures were altered from the PCC to the PCC-R, as noted above. Conversation with a child is tape-recorded to give a natural and representative speech sample. Around 3 minutes of continuous child speech is recorded using narrow phonetic transcription, omitting lengthy adult contributions and long silences. Knowledge of IPA symbols is required, as are phonetic transcription skills.

[136] The unit of analysis is consonants in a word, and the child's words are 'glossed' to give the adult version and so the correct consonant pattern. Only words that can be transcribed and glossed are scored. Although not part of Shriberg's procedures,[33] the PCC-R is in some assessments computed from single word elicitations checked against brief continuous speech samples as it can be difficult to gloss unintelligible continuous speech (see DEAP, below).

Target consonants are considered incorrect when they are either omitted, substituted by another consonant or glottal stop, inappropriately voiced or have extra phonemes added where these errors are not accounted for by dialectic variation or coarticulations associated with rapid speech. For the PCC, the response definition for children who obviously have speech errors is 'score as incorrect unless heard as correct',[33] but for the PCC-R, as noted, subphonemic distortions are counted as correct. Consonants in second or successive repetitions of an adjacent word or syllable are not scored unless they alter only the intended (adult-language) consonants in the first attempt. Calculation is a simple percentage: number of consonants correct/number of correct + incorrect consonants × 100. Severity ratings associated with percentage scores are: >85%, normal–mild; 65% to 85%, mild–moderate; 50% to 65%, moderate–severe; <50%, severe (*sic*: p. 115).[34]

Psychometric properties

In studies by the Shriberg and colleagues,[32–35] the PCC-R averaged a standard error of measurement (SEM) of 2.4 from 33 speech samples of children and adults across all consonants: 1.7 for the eight early developing consonants, 2.9 for the next eight to develop and 5.7 for the later eight. The increase in SEM was ascribed to transcription difficulties for later-developing consonants on which distortion errors may still be evident.[23] Inter-rater reliability measures showed that broad phonemic consonant transcription as used in the PCC-R achieved 92.7% agreement between two experienced transcribers.[35] The authors concluded that the relative gain in transcription accuracy obtained by the PCC-R gave reliability measures that were adequate for clinical and research purposes (p. 718). The PCC-R was considered to be the most appropriate metric from a list of alternative ways of counting phoneme realizations for comparison among 3- to 8-year-old children (p. 731),[35] giving sensitivity to true involvement through focusing on phonemic rather than subphonemic errors. The PCC-R distinguished children with typical speech from children with speech delays with no overlap, suggesting construct validity.

Further refinement of the PCC-R was undertaken by another research team using speech samples of typically developing children to compute a monthly performance growth curve for children from 18 to 172 months of age.[36] The aim was to track rapidly developing speech changes after childhood traumatic brain injury, and the initial results were successful, thus suggesting that high responsivity may be obtained.

The Diagnostic Evaluation of Articulation and Phonology

Overview and purpose

The Diagnostic Evaluation of Articulation and Phonology (DEAP)[37] aims both to identify subtypes of speech impairment by comparing a child's speech with developmental norms and to measure severity. Subtypes distinguish children using phoneme error patterns observed in typically developing children, albeit delayed or with developmentally early and later errors co-occurring, from children using developmentally unusual processes. A second distinction is between children whose error patterns are

Diagnostic Evaluation of Articulation and Phonology (DEAP)		
Purpose	To assess child speech production, differentially diagnose types of speech disorder and assess severity	
Population	Ages 3y–6y 11mo with speech delay or disorder or who are in the process of diagnosis.=	
Description of domains (subscales)	One domain/five subscales. Diagnostic Screen (10 pictures named twice; imitation of error phonemes); Articulation (30 pictures named; CV/VC syllable imitation); Oromotor [diadochokinetic (DDK) rates; isolated/sequenced oromotor movements]; Phonology (50 pictures named; target words in connected speech); Inconsistency (25 pictures named three times)	
Administration and test format	Time to complete: diagnostic screen 5min; full battery 30–40min Testing format: diagnostic screen then relevant subscales. Child names/describes stimulus pictures and carries out oral movements. Speech–language pathologist (SLP) transcribes using Institutional Phonetic Alphabet (IPA) symbols Scoring: standard scores/centile ranks for consonant inventory, DDK, isolated/sequenced oral movements, Percentage Consonant Correct (PCC), Percent Vowels Correct, Percent Phenomes Correct, single words vs connected speech and inconsistency. Qualitative analysis of typical and unusual error patterns Training: designed for qualified SLPs. Requires experience in speech transcription using IPA conventions. Restricted to purchasers with a formal qualification relevant to speech assessment	
Psychometric properties	Normative Sample: UK and Australia – 828 children, balanced for age, socioeconomic status and sex. US edition – 650 children based on current US population. Also, 83 bilingual children with English as a second language and Punjabi/Mirpuri/Urdu as a first (separate analyses provided) *Reliability* Internal consistency: not applicable – complete range of consonants assessed. Test–retest – quantitative subscales Pearson correlations 0.666 (sequenced oral movement) – 0.939 (PCC). Production of consonants 87.50–100% agreement, error patterns 98.21–100%. Inter-rater – quantitative subscales Pearson correlations 0.315 (Percent Vowels Correct)–0.886 (Percent Phenomes Correct). Production of consonants and error patterns 94.2–100% agreement *Validity* Content validity: complete coverage of the phonological system of English language Construct validity: high correlation (r=0.95; p<0.001) between DEAP-PCC and EAT Responsiveness: case studies show DEAP may detect changes in phoneme realisation over short periods	[140]
How to order	Pearson Cost: US$259	[141]
Key references	Dodd B, Hua Z, Crosbie S, Holm A, Ozanne A (2002) *Manual of Diagnostic Evaluation of Articulation and Phonology (DEAP)*, US edition. London: Psychological Corporation.	

consistent across all instances of a word pronounced in the same linguistic context, and those who are inconsistent. These distinctions have significance for intervention. The DEAP is for children aged 3 years, birth to 6 years, and 11 years. Examples of a child undertaking DEAP subtests may be downloaded from http://informahealthcare.com/doi/full/10.1080/14417040600861086.

Administration and scoring [137]

The 'Diagnostic Screen' identifies which aspects of speech need further investigation using three tasks: (1) 'naming'

10 pictures representing words containing the most single English consonants and some consonant clusters and vowels; (2) 'imitating' any phonemes that show errors on the naming task; and (3) 're-naming' the 10 screening words to check for consistency. The SLP transcribes responses live using broad phonemic IPA symbols.

If the child fails to imitate error phonemes appropriate to their age, an 'Articulation' assessment is made of the vowels and consonants of English using a picture-naming task, with any incorrect productions later imitated, and an 'Oromotor' assessment is made of non-speech movements. If the child makes consonant errors but can correctly imitate the relevant speech sounds, a 'Phonological' assessment analyses error patterns in both single words and connected speech, again using picture elicitation and distinguishing developmentally typical from developmentally unusual errors. Where half or more words are produced differently on the two screening trials, a further 'Inconsistency' assessment uses a 25-word naming task to establish consistency of
[139] production, which is repeated three times with distracting activities between trials, and again the 'Oromotor' assessment is undertaken. The diagnostic screen takes around 5 minutes, and the full assessment battery around 30 to 40 minutes.[38]

The unit of analysis is the syllable for the 'Phonological' assessment, the word for the 'Inconsistency' assessment and the phoneme for the 'Articulation' assessment. Picture and task materials for each assessment are included in the package. Standard scores/centile ranks are given for the consonant 'inventory' (the list of consonants used in speech at some time), 'diadochokinetic' rates, 'isolated oral movements', 'sequenced oral movements', PCC; 'Percent Vowels Correct', 'Percent Phonemes Correct' (consonants plus vowels), 'single words versus connected speech' and 'inconsistency'. Responses are recorded and scored on record forms. Qualitative analysis of error patterns compares the child with children of their own age, distinguishing among errors that would be used by at least 10% of typically developing children in the age band, delayed errors that would be used by at least 10% of younger children and unusual error patterns that would not be used by more than 10% of children in the normative sample at any age.

Psychometric properties

Information is taken from the DEAP manual.[37] A total of 1478 children were included in the standardization
[138] sample, with 828 balanced for age, socioeconomic status and sex from the UK and Australia and 650 in a balanced US sample. Test–retest reliability for quantitative measures for 56 children showed Pearson's correlations of 0.67 (sequenced movement) to 0.94 (PCC). Per cent agreement on consonant production ranged from 87.50% ('–l') to 100% (13 consonants). Error patterns ranged from 98.21% to 100% agreement. Inter-rater reliability from 69 children and two judges showed Pearson's correlations from 0.32 (per cent vowels correct) to 0.89 (PCC). Consonant production and error patterns ranged from 94.2% to 100%. Sensitivity trials on a total of 57 age-appropriate children compared the Diagnostic Screen with the full DEAP assessment, giving 10% to 13% false positives, but no false negatives. Specificity assessed how accurately the Diagnostic Screen identified which further DEAP tests to use and gave 93% agreement between two SLPs blinded to DEAP test results.

Content validity was established by the complete coverage of the phonological system of the English language, and concurrent validity by comparing 50 children on the Edinburgh Articulation Test,[39] which showed a high correlation (r=0.95; p<0.001) between PCC measured by the DEAP and the EAT. Responsiveness is suggested by case study data.

Fluency and Rhythm of Speech Functions – Stuttering (b330)

Percentage of Syllables Stuttered

Overview and purpose

Non-fluency is a feature of the speech of young children, and distinguishing normal non-fluency from stuttering can be problematic, with difficulties in collecting representative speech data and in establishing inter-rater reliability. However, a measured reduction in the number of moments of overt stuttering is a frequent intervention goal for young children. Percentage of Syllables Stuttered (PSS)[40] has been developed as a core measure of stuttering frequency to deal with these known measurement difficulties. The measure is the percentage of syllables unambiguously associated with stuttering in a sample of child speech, with a reduction in the PSS used as an outcome measure.

Administration and scoring

An adult engages the child in naturalistic conversation: at least 300 child syllables are needed for analysis.[40] A trained SLP unobtrusively presses one of two buttons on a commercially available tally counter as the child speaks: one button for every syllable judged as free of stuttering, the other for stuttered syllables. Such online judgements require training, but tape-recorded samples may be used.

Percentage of Syllables Stuttered (PSS)	
Purpose	To assess core stuttering behaviour and severity
Population	Children of any age who stutter or who are in the process of diagnosis
Description of domains (subscales)	One domain/no subscales
Administration and test format	Time to complete: around 5min Testing format: around 5min of child speech (at least 300 syllables) is recorded Scoring: speech–language pathologist (SLP) observes and codes syllables as showing unambiguous stammering or not, live, using a push-button recorder Training: designed for SLP use. No training, but practice in accurate recording needed
Psychometric properties	Normative sample: no normative sample *Reliability* Internal consistency: averages across syllables. Test–retest – variation anticipated across samples Inter-rater: around 75% of instances of child stuttering/not stuttering agreed by 80% of highly experienced judges. Inter- and intrarater intraclass correlation coefficient 0.99, with experienced research raters *Validity* Content validity: dichotomous direct measure. Construct/discriminant validity – Spearman's correlation 0.91, with severity rating by adults. Case data show that PSS varies along with parental measures of severity Responsiveness: measured changes between pre-intervention and 1-year postintervention follow-up
How to order	Published in book form as Onslow M, Packman A, Harrison E (2003) *The Lidcombe Programme of Early Stuttering Intervention*. Austin, TX: Pro-Ed Publishing.
Key references	Bothe AK (2008) Identification of children's stuttered and nonstuttered speech by highly experienced judges: binary judgements and comparisons with disfluency-types definitions. *J Speech Lang Hear Res* 51: 867–878.

Where measurement is carried out online, the time taken is the few minutes needed to collect a sample of 300 syllables from a child. There is no age limit. The score is a simple percentage of stuttered syllables divided by stuttered plus unstuttered syllables.

Psychometric properties

Test–retest judgements are not appropriate, as the percentage of stuttering is expected to vary across samples. However, samples are taken and scored frequently and PSS is expected to decrease over time during intervention. Inter-rater measures of identifying stuttering have shown that 80% of highly experienced judges agreed on the presence or absence of stuttering in approximately three-quarters of brief samples from children, aged 2 to 8 years, who stuttered.[41] Trained and experienced SLPs showed intraclass correlations of r=0.99 for both intra- and interjudge reliability;[42] and the average differences between a researcher's pretreatment PSS measurement and an independent experienced blinded observer were 2.3 PSS, with no differences after treatment.[43] Measuring PSS reliably is therefore possible with training, but inexperienced users should test their standards of intra- and preferably inter-rater reliability before using the measure.

Comparison between the PSS and a nine-point severity scale showed a Spearman correlation of 0.91 for adults who stuttered, suggesting construct validity,[44] and case study data showed PSS reducing over time in line with parental judgements of severity.[40] Responsivity was shown by reductions in PSS from pre-intervention to 1-year follow-up, with interventions of varied durations.[43]

The Stuttering Severity Instrument for Children and Adults

Overview and purpose

The Stutterng Severity Instrument for Children and Adults (SSI-4)[45] instrument was first published in 1972[45] and is now in its fourth edition. It provides a standardized assessment of the quality and quantity of dysfluency for children aged 2 to 10 years and older. The subscales are 'frequency', 'duration', observable 'physical concomitants of stuttering' and a speech 'naturalness' scale.

Administration and scoring

Assessment takes 15 to 20 minutes. Pictures and conversation elicit child talk, which is tape recorded. At least 150 child words are needed. 'Frequency' of dysfluency is measured by the PSS. 'Duration' is scaled using the average length of the three longest stuttering moments, measured to one-tenth of a second. Four observable distracting 'physical concomitants' of stuttering are scored: sounds which accompany stuttering, such as throat clearing; facial grimaces, such as eye blinks or tongue protrusion; head movements, such as head turning to avoid eye contact; and arm or leg movements, such as foot tapping. Scores are expressed as scale scores, and a total score may be calculated. There is the option of automatic computerized scoring of frequency and duration.

Psychometric properties

The normative sample for the SSI-4 was 72 preschool children, 139 school-aged children and 60 adults. The third edition of the instrument (SSI-3) showed an average inter-rater difference between a pretreatment measurement and a blinded expert comparison of 2.25 units, with no rater differences after treatment.[43] The SSI-3 was reviewed in HSTAT 52,[15] but did not meet relaxed validity or reliability criteria. SSI-3 scores showed change from pre-intervention to 1-year follow-up after intervention, with interventions varying in duration.[43]

REFERENCES

1. Ma EPM, Yiu EML, Abbott KV (2007) Application of the ICF in voice disorders. *Semin Speech Lang* 28: 343–350.
2. McLeod S, McCormack J (2007) Application of the ICF and ICF-children and youth in children with speech impairment. *Semin Speech Lang* 28: 254–264.
3. Yaruss JS, Quesal RW (2004) Stuttering and the International Classification of Functioning, Disability, and Health: an update. *J Commun Disord* 37: 35–52.
4. World Health Organization (2006) *International Classification of Diseases and Related Health Problems*, 10th edition. Geneva: World Health Organization.
5. Venail F, Gardiner Q, Mondain M (2004) ENT and speech disorders in children with Down's syndrome: an overview of pathophysiology, clinical features, treatments and current management. *Clin Pediatr* 43: 783–790.
6. Albertini G, Bonassi S, Dall'Armi V Giachetti I, Giaquinto S, Mignano M (2010) Spectral analysis of the voice in Down syndrome. *Res Dev Disabil* 31: 995–1001.
7. McLeod S, Harrison LJ (2009) Epidemiology of speech and language impairment in a nationally representative sample of 4- to 5-year old children. *J Speech Lang Hearing Res* 52: 1213–1229.
8. McKinnon DH, McLeod S, Reilly S (2007) The prevalence of stuttering, voice, and speech-sound disorders in primary school students in Australia. *Lang Speech Hearing Services Schools* 38: 5–15.
9. Broomfield J, Dodd B (2004) The nature of referred subtypes of primary speech disability. *Child Lang Teach Ther* 20: 135–151.
10. McLeod S (2006) An holistic view of a child with unintelligible speech: insights from the ICF and ICF-CY. *Adv Speech Lang Pathol* 8: 293–315.
11. Walsh R (2005) Meaning and purpose: a conceptual model for speech pathology terminology. *Int J Speech Lang Pathol* 7: 65–76.
12. Holm A, Crosbie S, Dodd B (2007) Differentiating normal variability from inconsistency in children's speech: normative data. *Int J Lang Commun Disord* 42: 467–486.
13. Broomfield J, Dodd B (2004) The nature of referred subtypes of primary speech disability. *Child Lang Teach Ther* 20: 135–151.
14. Yaruss JS (1997) Clinical implications of situational variability in preschool children who stutter. *J Fluency Disord* 22: 187–203.
15. Biddle A, Watson L, Hooper C, Lohr KN, Sutton SF (2002) *Evidence/Report Technology Assessment No. 52*. (Prepared by the University of North Carolina Evidence-based Practice Center under Contract No 290–97–0011). AHRQ Publication No. 02-E010. Rockville, MD: Agency for Healthcare Research and Quality.
16. Consensus Auditory-Perceptual Evaluation of voice (CAPE-V) (2009) *ASHA Special Interest Division 3, Voice and Voice Disorders*. Available at: www.asha.org/NR/rdonlyres/C6E5F616–972F-445A-AA40–7936BB49FCE3/0/D3CAPEVprocedures.pdf (accessed 3 June 2009).
17. Kempster GB, Gerratt BR, Verdolini AK, Barkmeier-Kraemer J, Hillman RE (2009) Consensus Auditory-Perceptual Evaluation of Voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol/Am Speech Lang Hearing Assoc* 18: 124–132.
18. Braden MN, Johns MM, Klein AM, Delguadio JM, Gilman M, Hapner ER (2010) Assessing the effectiveness of Botulinum toxin injections for adductor spasmodic dysphonia: clinician and patient perception. *J Voice* 20: 242–249.
19. Eadie TL, Baylor CR (2006) The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *J Voice* 20: 527–544.
20. Oates J (2009) Auditory-perceptual evaluation of disordered voice quality: pros, cons and future directions. *Folia Phoniatrica Logopaedica* 61: 49–56.

21. Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA, Hoffman HT (2007) Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice* 21: 576–590.
22. Awan SN, Lawson LL (2009) The effects of anchor modality on the reliability of vocal severity ratings. *J Voice* 23: 341–352.
23. Kelchner LN, Brehm SB, Weinrich B, et al (2009) Perceptual evaluation of severe pediatric voice disorders: rater reliability using the consensus auditory perceptual evaluation of voice. *J Voice*.
24. Hirano M (1981) Psycho-acoustic evaluation of voice. In: Arnold W, Wyke W, editors. *Disorders of Human Communication 5, Clinical Examination of Voice*. New York and Wien: Springer-Verlag, pp. 81–84.
25. Lee MD, Milstein C, Hicks D, Akst LM, Esclamado RM (2007) Results of ansa to recurrent laryngeal nerve reinnervation. *Otolaryngol Head Neck Surg* 136: 450–454.
26. Boone DR (1980) *The Boone Voice Program for Children: Screening, Evaluation and Referral Manual*, 2nd edition. Austin, TX: Pro-Ed.
27. Colton RH, Casper JK, Leonard R (2006) *Understanding Voice Problems*. Philadelphia: Lippincott Williams & Wilkins.
28. Fendler M, Shearer WM (1988) Reliability of the S/Z Ratio in normal children's voices. *Lang Speech Hearing Serv Schools* 19: 2–4.
29. Eckel FC, Boone DR (1981) The s/z ratio as an indicator of laryngeal pathology. *ASHA* 46: 147–149.
30. Hufnagle J, Hufnagle KK (1988) S/Z ratio in dysphonic children with and without vocal cord nodules. *Lang Speech Hearing Serv Schools* 19: 418–422.
31. Shriberg LD, Austin D, Lewis BA, McSweeny JL, Wilson DL (1997) The percentage of consonants correct (PCC) metric: extensions and reliability data. *J Speech Lang Hearing Res* 40: 708–722.
32. Shriberg LD, Kwiatkowski J (1982) Phonological disorders III: a procedure for assessing severity of involvement. *J Speech Hearing Disord* 47: 256–270.
33. Shriberg LD (1993) Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *J Speech Lang Hearing Res* 36: 105–140.
34. Shriberg LD, Austin D, Lewis BA, McSweeny JL, Wilson DL (1997) The speech disorders classification system (SDCS): extensions and lifespan reference data. *J Speech Lang Hearing Res* 40: 723–740.
35. Campbell TF, Dollaghan C, Janosky JE, Adelson PD (2008) A performance curve for assessing change in Percentage of Consonants Correct Revised (PCC-R). *J Speech Lang Hearing Res* 50: 1110–1119. [Erratum appears in *J Speech Lang Hear Res* (2008) 51: 301] [142]
36. Dodd B, Hua Z, Crosbie S, Holm A, Ozanne A (2002) *Manual of Diagnostic Evaluation of Articulation and Phonology (DEAP)*. London: Psychological Corporation.
37. Leahy M, Dodd B. *DEAP test standardization*. Available at: www.tcd.ie/slscs/csls/research/featuredresearch_deap_standardisation.php (accessed 4 June 2009). [143]
38. Anthony A, Bogle D, Ingram TTS, McIsaac MW (1971) *Edinburgh Articulation Test*. Edinburgh: Churchill Livingstone.
39. Onslow M, Packman A, Harrison E (2003) *The Lidcombe Program of Early Stuttering Intervention*. Austin, TX: Pro-Ed Publishing, pp. 59–63.
40. Bothe AK (2008) Identification of children's stuttered and nonstuttered speech by highly experienced judges: binary judgments and comparisons with disfluency-types definitions. *J Speech Lang Hear Res* 51: 867–878.
41. Jones M, Onslow M, Packman A, et al (2005) Randomised controlled trial of the Lidcombe Program of early stuttering intervention. *BMJ* 331(7518): 659.
42. Miller B, Guitar B (2009) Long-term outcome of the Lidcombe Program for early stuttering intervention. *Am J Speech Lang Pathol* 18: 42–49.
43. O'Brian S, Packman A, Onslow M, O'Brian N (2004) Measurement of stuttering in adults: comparison of stuttering-rate and severity-scaling methods. *J Speech Lang Hear Res* 47: 1081–1087.
44. Riley GD (2008) *Stuttering Severity Instrument for Children and Adults*, 4th edition. East Moline, IL: LinguiSystems.