ABSTRACT: **Purpose:** The aims of this review were to identify and describe the assessment tools that have been developed to assess prosodic skills in children and adults and to evaluate the clinical utility of the tools. **Method:** Currently available tools were identified through searching 4 online databases and bibliographies of relevant articles and by contacting authors. **Results:** Nine assessment tools were identified. The tools were appraised for their intended purpose, target population, domains of prosody assessed, validity, reliability, and normative sample data. The purpose of development and the content of the tools were well documented, but data on how feasible they were to use in practice were scarce. Each tool met some but not all of the widely accepted criteria for validity and reliability. Most have not been sufficiently well tested for use in routine clinical practice. **Conclusion:** This review highlights the need to continue to develop and test tools for the effective and comprehensive assessment of prosodic skills in children and adults.

**KEY WORDS:** PEPS–C, DANVA 2, FAB, MNTAP, Aprosodia Battery, ACS, PPAT
on the part of clinicians in making efforts to assess prosody in client populations.

Following Crystal (2009), we found that there are no published papers that have systematically evaluated and compiled the various tools that are available to assess prosodic skills in children and adults. It is surprising that this lack of literature on assessment tools exists despite the fact that prosodic difficulties extend across a wide range of communication disorders such as autism spectrum disorder (ASD; Green & Tobin, 2009; McCann & Peppé, 2003), specific language impairment (SLI; Marshall, Harcourt-Brown, Ramus, & van der Lely, 2009; Stojanovik, Setter, & Ewijk, 2007), Parkinson’s disease (Martens et al., 2011), apraxia (Odell & Shriberg, 2001), aphasia (Danly & Shapiro, 1982), brain injury (Karow, Marquardt, & Marshall, 2001; Moen, 2009; Ross, Edmondson, Seibert, & Homan, 1988), and hearing loss (Nakata, Trehub, & Kanda, 2012; Peng, Tomblin, & Turner, 2008).

Theoretical approaches to explaining prosody (phonetic and phonological perspectives) and the differences in prosody error profile across clinical populations may have contributed to the development of different assessment approaches. Pitch, duration, and stress are the phonetic correlates of prosody. The physical correlates of these features are the speech’s fundamental frequency ($F_0$), syllable duration, and intensity, respectively. Phonological correlates of prosody include the variations in pitch, length, and loudness that are produced in speech for conveying subtle changes in the meaning of spoken messages independent of words and grammatical order (Roach, 2000). In other words, stress and intonation are used to convey the grammatical, affective, and pragmatic functions of language.

Methods for assessing disordered prosody can be classified into instrumental approaches, including software applications, and approaches to measure prosodic functions. Instrumental approaches involving acoustic analysis (e.g., PRAAT, Boersma, & Weenink, 2001; EMU, Bombien, Cassidy, Harrington, John, & Palethorpe, 2006) focus on displaying and quantifying the relevant acoustic correlates of prosody (Shriberg, Kwiatkowski, Rasmussen, Lof, & Miller, 1992), whereas prosodic function measures assess the phonological or communicative aspects of prosody. The autosegmental metrical framework for intonational analysis (Pierrehumbert, 1980) and the related transcription systems such as ToBI and IViE (Grabe, Nolan, & Farrar, 1998) provide options for prosodic labeling.

Earlier studies on communication disorders have evaluated prosody in terms of its acoustic dimensions. More recently, emphasizing the role of prosody in communicative efficiency, the assessment of functions of prosody has been advocated rather than quantifying $F_0$, duration, and intensity parameters (e.g., Profiling Elements of Prosody in Speech-Communication [PEPS–C], Peppé & McCann, 2003; Diagnostic Analysis of Nonverbal Accuracy 2 [DANVA 2], Nowicki & Duke, 1994). The choice of instrumental or prosodic function measures to assess prosody should depend on the prosodic difficulties of the clinical population being assessed. Several studies have acoustically analyzed echolalia, imitative and spontaneous speech in conversation, and narratives in children, adolescents, and adults with ASD and have found important prosodic differences in the variance of $F_0$, duration of syllables; use of prosodic contours (Diehl & Paul, 2009; Green & Tobin, 2009); and coordination of prosodic cues such as pitch, duration, and amplitude (Van Santen, Prud’hommeaux, Black, & Mitchell, 2010).

Diehl and Paul (2013) conducted acoustic and perceptual measurements of prosody production by children with ASD and reported that differences in acoustic parameters were present in the speech of the group with ASD even when the different aspects of prosody were perceived accurately. Inaccurate production of acoustic features, such as excessive or misaligned pitch, slow syllable-timed speech, fast rate of speech, and monoloudness, have been reported in individuals with apraxia of speech (Barry, 1995), PD (Penneker, Miller, Hertrich, Ackermann, & Schumann, 2001; Ma, Whitehill, & Cheung, 2010) and other communication disorders (Shriberg et al., 2001). Darley, Aronson, and Brown (1969, 1975) reported that prosodic deficits play a significant role in the characterization of motor speech disorders. Prosodic deficits were also reported as a core feature of childhood apraxia of speech by the American Speech-Language-Hearing Association (ASHA, 2007). In general, inaccurate prosody productions affect an individual’s speech intelligibility (Chin, Bergeson, & Phan, 2012; Klopfenstein, 2009; Mayo, Aubanel, & Cooke, 2012).

This review focuses on assessment tools that are used to probe for the communicative aspects of prosody. The functions of prosody are identified at the indexical, grammatical, affective, and pragmatic levels of communication. Indexical information is information that is related to the age, identity, and gender of the speaker (Paul, Augustyn, Klin, & Volkmar, 2005; Romero-Trillo & Newell, 2012; Stojanovik, 2010). The grammatical functions of prosody include (a) determining the boundaries of phrases, clauses, or sentences, particularly when there is ambiguity (e.g., /FRUIT, SALAD and MILK/ vs. /FRUIT-SALAD and MILK/; Wells & Peppé, 2003) and (b) differentiating between word classes when there are homonyms.
The **affective** function of prosody is the use of prosodic features such as pitch contour, pauses, and word stress to express the speaker’s emotions and attitudes (Peppé, 2009; Roach, 2000). Prosodic patterns convey different emotions (Banse & Scherer, 1996; Juslin & Laukka, 2003). Happiness, for example, is characterized by a fast speaking rate, rising pitch, high variability, and fast voice onsets, and sadness is nearly the opposite (Hirschberg, 2002). The use of prosodic patterns to convey discourse functions is the **pragmatic** aspect of prosody. Prosody is used to signal to the listener what is to be taken as **new** information and what is already **given**. Prosody also helps the listener to distinguish between questions, statements, and commands (Roach, 2000). A variety of prosodic features are used by speakers to indicate to others that they have finished speaking, that another person is expected to speak, that a particular type of response is required, and so on. For example, a rising tone at the end of an utterance typically indicates that a response is required, and a falling tone suggests the end of a conversation (McCann & Peppé, 2003).

It is important to note that researchers have reported the overlapping functions of these different aspects of prosody. For example, the use of prosody to distinguish between question and statement was classified as a grammatical function by Paul et al. (2005), whereas Wells and Peppé (2003) described it as the pragmatic aspect of prosody. Diehl and Paul (2009) reported that the use of prosody to distinguish between question and statement can fit into both categories as it conveys the sentence type and also signals a type of mental state or intent of discourse (either the end of a conversation or a particular response is required).

Although there are differences in classification, it is widely recognized that the accurate perception and production of the ranging aspects of prosody are a significant component of successful social communication (Aziz-Zadeh, Sheng, & Gheytanchi, 2010). Difficulties in perceiving the communicative aspects of prosody as well as deficits in prosody productions have been reported in children with hearing loss by several researchers (Meister, Landwehr, Pyschny, Walger, and von Wedel, 2009; Nakata et al., 2012; Peng et al., 2008). Hence, the combined use of instrumental techniques and prosodic function measures to assess prosodic skills is advised for this population.

Certain aspects of prosody are more relevant than others in specific client groups. For example, prosody evaluation in adults with neurological and psychiatric impairments has mainly focused on investigating the perception and production of affective prosodic skills (Moen, 2009; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009), whereas the evaluation of grammatical, affective, and pragmatic aspects of prosody is relevant in individuals with hearing loss, ASD, and SLI. The important communicative functions of prosody and evidence for prosodic deficits in various speech-language and hearing disorders makes it important to assess prosodic skills in both typical and clinical populations.

Diehl and Paul (2009) compared three prosody assessment measures—the Prosody Profile (PROP; Crystal, 1982), Prosody-Voice Screening Profile (PVSP; Shriberg, Kwiatkowski, & Rasmussen, 1990), and PEPS-C—to the other well-established methods of assessing language, such as the Peabody Picture Vocabulary Test—Fourth Edition (PPVT–IV; Dunn & Dunn, 2007) and the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF–IV; Semel, Wiig, & Secord, 2003). Diehl and Paul reported that there are no analogous data on the typical developmental sequence of prosody acquisition and adequate psychometrics derived from studies of spontaneous language use. Diehl and Paul also reported that there is a need for a prosody assessment tool that (a) has a representative normative sample and strong psychometric properties, (b) is based on empirical information regarding the typical sequence of prosodic acquisition and is developmentally sensitive, (c) assesses various domains of prosody, (d) uses tasks that have high ecological validity, and (e) has established clinical utility.

We decided to review the available published assessment tools that are used to assess prosodic skills in children and adults, explaining the different domains of prosody that can be assessed using these tools, and providing information on the normative data and psychometric properties of each tool and its clinical utility. This review provides clinicians and researchers with information that will enable them to select the appropriate tool(s) for the assessment of prosodic skills in children and adults.

**METHOD**

**Literature Search Strategy**

We conducted a review of literature to identify clinical and research tools that have been developed so far to assess prosodic skills in children and adults. The goal was to identify tools with established reliability, validity, and normative data. We entered the search terms in Table 1 in different combinations into four online databases: ScienceDirect, SCOPUS,
Web of Science, and PsycINFO. We also conducted manual searches of the bibliographies of published reviews and articles and contacted test developers to gain additional information regarding the tools. Inter-library loan facilities at the University of Auckland were used to access unpublished dissertation work.

**Assessment Tool Inclusion/Exclusion Criteria**

We included assessment tools that were published in English and that were developed for use with individuals with acquired neurological disorders and psychiatric disorders, even though these tend to be components of a larger test battery and thus are less comprehensive than the larger battery. Tools that were translated from English to other languages or were developed in languages other than English were excluded from this review (e.g., Foley, Gibbon, & Peppé, 2011; Ladani et al., 2012; Martínez-Castilla & Peppé, 2008; Torppa et al., 2014; Van Zyl & Hanekom, 2013). Assessment tools that were out of print, such as the Right Hemisphere Language Battery (Bryan, 1989), which includes a subtest to assess the production of emphatic stress, are not described in this review.

**Evaluation Criteria**

We appraised each assessment tool against a range of criteria pertinent to whether a tool might be suitable for clinical use. The clinical utility of each tool was gauged through examining features such as intended purpose, target population, domains of prosody assessed, validity and reliability data, nature of the normative sample, time and ease of administration and scoring, availability in different formats, and administration method.

**RESULTS AND DISCUSSION**

We identified nine assessment tools: PROP, PVSP, PEPS–C, Perception of Prosody Assessment Tool (PPAT; Klieve, 1998), Minnesota Tests of Affective Processing (MNTAP; Lai, Hughes, & Shapiro, 1991), DANVA 2, Aprosodia Battery (Ross, Thompson, & Yenkoshy, 1997), Florida Affect Battery (FAB; Bowers, Blonder, & Heilman, 1999), and Advanced Clinical Solutions (ACS; Pearson, 2009). We are not aware of any additional batteries that assess prosody.

We divided the assessment tools into two sections based on the number of aspects of prosody that they assess. Table 2 shows tools with subtests that assess two or more aspects of prosody; Table 3 shows tools that involve only affective prosody. Tables 2 and 3 also summarize each tool’s target population, subtests involved, domains of prosody assessed, normative sample characteristics, and reliability and validity information available either in the literature or from the test developer. The tools differ in terms of target population, normative data, psychometric properties, and the domains of prosody assessed. The feasibility of each assessment tool is discussed based on factors such as time taken to administer, test format, appropriateness of test items, and ease of scoring.

**Intended Purpose and Target Population**

Tables 2 and 3 indicate the different purposes for which the tools were developed. These include a focus on assessing (a) expressive prosodic skills, (b) perception of affective prosody in adults with neurologic and psychiatric disorders, (c) social and cognitive functioning in adults, and (d) prosodic skills in children with ASD and hearing loss. There are differences in the conceptual frameworks on which these tools are based: psycholinguistic (PEPS–C, PPAT), neurolinguistic (Aprosodia Battery), and social and cognitive functioning (DANVA 2, MNTAP, FAB, ACS), reflecting the many identified roles of prosody in social communication.

The PEPS–C has been used in a number of research studies investigating prosodic skills in typically developing children and children with communication disorders (Catterall, Howard, Stojanovik, Szczerbinski, & Wells, 2006; Foley et al., 2011; Martínez-Castilla & Peppé, 2008; Peppé, & McCann, 2003; Peppé, McCann, Gibbon, O’Hare, &
<table>
<thead>
<tr>
<th>Test</th>
<th>Age (in years)</th>
<th>Purpose and population</th>
<th>Subtests</th>
<th>Prosody subtests receptive/expressive</th>
<th>Aspects of prosody</th>
<th>Normative sample and inclusion criteria</th>
<th>Reliability/Validity</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROP</td>
<td>Children &amp; adults</td>
<td>To obtain information about the expressive prosodic patterns encountered in a sample of clinical data.</td>
<td>Four prosodic patterns</td>
<td>Receptive: None</td>
<td>Pitch, tempo, stress</td>
<td>No normative, reliability &amp; validity data provided. Age ranges, guidelines &amp; examples of impaired prosody are provided.</td>
<td>Not available.</td>
<td>Time: depends on expertise in transcription</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Expressive: Intonation (nuclear pitch direction). Tempo (phrasing), Stress (phrasal) &amp; strategies for producing stress</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PVSP</td>
<td>3–81</td>
<td>To assess speakers' prosody and voice in conversational speech.</td>
<td>7 supra-segmentals (3 prosody, 4 voice)</td>
<td>Receptive: None</td>
<td>Phrasing, rate, pitch, stress, loudness, laryngeal quality and resonance</td>
<td>252 audiotaped exemplars from 3- to 19-year-old children with normal &amp; disordered speech development were selected &amp; coded using Dictaphone 2550 audiocassette playback device. CSpeech by Milenkovic (1991) &amp; VOCAL (Milenkovic, 1989) were used for instrumental validity study.</td>
<td>Perceptual criterion validity for pitch, quality &amp; resonance ranged between 71%–84%. Instrumental criterion validity for rate, stress, pitch &amp; quality ranged between 80%–100%.</td>
<td>Time: depends on expertise in transcription</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEPS–C</td>
<td>5–14</td>
<td>To assess receptive and expressive prosodic skills in children. Widely used in children with autism spectrum disorder.</td>
<td>12 subtests (6 receptive prosody &amp; 6 expressive prosody)</td>
<td>Receptive: Short Item Discrimination, Turn-End Reception, Turn-End Expression, Affect Reception, Affect Expression, Long Item Discrimination, Long Item Imitation, Chunking Reception, Chunking Expression, Contrastive Stress Reception, Contrastive Stress Expression</td>
<td>Pitch direction, grammatical, affective (like, dislike), pragmatic</td>
<td>Nomative data on 120 students from North London ages 5–14 years. English as first language; no identified speech &amp; language problems or educational problems, &amp; residents of England for at least 3 years. Normative data including mean &amp; SDs for ages 5:0 (years; months), 8:0, 10:0, and 13:0 are provided for each subtest (Wells, Peppé, &amp; Goulandris, 2004). The performance improved between the ages of 5:0 and 14:3.</td>
<td>Test–retest reliability for 30 participants with a 6-month interval were not significantly different, the range of variation +2.08 to –1.04. Intrarater reliability was checked by rescoring 18/30 participants (3-month interval); difference in scores was 2.6%. Interrater reliability was checked by obtaining 2</td>
<td>Time: –60 min</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2 (p. 1 of 2). Features of the assessment tools that evaluate two or more aspects of prosody.**
Table 2 (p. 2 of 2). Features of the assessment tools that evaluate two or more aspects of prosody.

<table>
<thead>
<tr>
<th>Test</th>
<th>Age  (in years)</th>
<th>Purpose and population</th>
<th>Subtests</th>
<th>Prosody subtests receptive/expressive</th>
<th>Aspects of prosody</th>
<th>Normative sample and inclusion criteria</th>
<th>Reliability/ Validity</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPAT (Klieve, 1998)</td>
<td>7–12</td>
<td>To evaluate prosodic perception in children using cochlear implants</td>
<td>Six subtests: Apart from linguistic context, Tone &amp; affect, Grammatical class, Stress, Compound and abutting words.</td>
<td>Expression, Contrastive Stress Expression</td>
<td>6 children ages 7–12 years. Attended oral school for children with HL, used Nucleus 22 multichannel CI with SPEAK strategy &amp; more than 1 year of experience with the implant. Participants varied in age of onset of HL, etiology of deafness, length of profound deafness pre implant &amp; experience with the device (Klieve &amp; Jeanes, 2001).</td>
<td>CI participants perceived prosodic cues of duration, intensity, &amp; pitch apart from a linguistic context above chance level, 70% or above. Performance on perceiving the prosodic cues meant within a linguistic context was at or just below chance level, between 60% and 70%.</td>
<td>Time: ~60 min Format: electronic administration; manual scoring Made for: research use</td>
<td></td>
</tr>
</tbody>
</table>

Note. PROP = Prosody Profile, PVSP = Prosody-Voice Screening Profile, PEPS–C = Profiling Elements of Prosody in Speech-Communication, PPAT = Perception of Prosody Assessment Tool.
### Table 3 (p. 1 of 3). Features of the assessment tools that evaluate affective prosody only.

<table>
<thead>
<tr>
<th>Test</th>
<th>Age (in years)</th>
<th>Purpose and population</th>
<th>Subtests</th>
<th>Prosody subtests receptive/expressive</th>
<th>Aspects of prosody</th>
<th>Normative sample and inclusion criteria</th>
<th>Reliability/Validity</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNTAP</td>
<td>6–11</td>
<td>To assess face perception and recognition of affective stimuli as conveyed through facial expression, language, and prosody.</td>
<td>16 subtests (4 auditory, 11 visual)</td>
<td>Receptive: Prosody/content preference, Lexical comprehension, Prosody/content congruence, Cross modal matching</td>
<td>Affective – happy, mad, sad, scared, neutral</td>
<td>67 children with ADHD and 38 controls ages 6–11. Inclusion criteria for ADHD group included 1) a teacher rating on the revised Conners teacher rating scale of 1.7 or greater on the hyperactivity index and 2) diagnosis of ADHD based on meeting 8 of 14 criteria set forth in the DSM III–R. No information on SES or ethnicity (Shapiro, Hughes, August, &amp; Bloomquist, 1993).</td>
<td>MNTAP scores for both groups were compared by sex &amp; correlated with age. No sex differences were found. Correlations with age were highest with Inverted faces, Face and Object recognition memory, &amp; Localization memory tasks. ADHD group differed from control group in tasks of prosody/content congruence &amp; cross-modal matching.</td>
<td>Time: ~2–3 hours</td>
</tr>
<tr>
<td>DANVA 2</td>
<td>3–99</td>
<td>To examine the perception of facial expression, paralanguage (emotional aspect of prosody), and understanding of body postures in children and adults.</td>
<td>5 subtests (2 faces, 2 paralanguage &amp; 1 posture)</td>
<td>Receptive: Adult paralanguage 2, Child paralanguage 2</td>
<td>Expressive: None</td>
<td>Reliability: Internal consistency DANVA 2 AF-158 college students; DANVA 2 CF-across 10 studies with children aged 4-16 years; DANVA 2 AP-M=33.5 years, N=20; DANVA 2 CP-8-year-old (N=32) &amp; 10-year-old (N=31); DANVA 2 adult postures- college students (N=54). Ethnicity comparable to community rates. Groups matched for SES &amp; IQ. Mean &amp; SDs of errors on adult &amp; child facial expressions &amp; paralanguage subtests are provided for ages 3–99. Data for child postures subtest-5–14 years; adult posture subtest-15–50 years.</td>
<td>Internal consistency: DANVA 2 AF–α = 0.90 DANVA 2 CF–α=0.69–0.81 DANVA 2 AP–α = 0.75 DANVA 2 CP–α = 0.74 (8 year old) &amp; α = 0.76 (10-year-old). DANVA 2 adult postures-α = 0.75.</td>
<td>Time: ~60 min</td>
</tr>
<tr>
<td>Test</td>
<td>Age (in years)</td>
<td>Purpose and population</td>
<td>Subtests</td>
<td>Prosody subtests (receptive/expressive)</td>
<td>Aspects of prosody</td>
<td>Normative sample and inclusion criteria</td>
<td>Reliability/Validity</td>
<td>Feasibility</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>----------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------------------------------------</td>
<td>------------------------------------------</td>
<td>---------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Aprosodia battery (Ross et al., 1997)</td>
<td>&gt;17</td>
<td>To examine reception and expression of affective prosody in adults with acquired neurological disorders.</td>
<td>Four subtests (2 receptive &amp; 2 expressive)</td>
<td>Receptive: Affective–prosodic comprehension (word, monosyllabic and asyllabic levels), Affective–prosodic discrimination (word level)</td>
<td>Affective– happy, sad, angry, surprised, neutral, disinterested</td>
<td>22 brain-damaged subjects and 16 control subjects mean ages 49.4± 15.2, M/F–9/7. Six weeks poststroke patients with unilateral hemispheric infarctions were included. Controls had no history of previous neurological problems, strongly right handed as determined by a score of +70 on the Edinburgh Inventory (Oldfield, 1971). Groups were matched for sex &amp; education (Ross et al., 1997).</td>
<td>Results confirmed that the mean coefficient of variation calculated is an appropriate measure of affective performance &amp; that the performance of control &gt; LHD &gt;&gt; RHD</td>
<td>Time: ~60 min Format: electronic administration; manual scoring Made for: research use</td>
</tr>
<tr>
<td>FAB (Bowers et al., 1999)</td>
<td>&gt;17</td>
<td>To assess perception of affective prosody in adults with neurologic or psychiatric disorders.</td>
<td>10 subtests (5 facial, 3 prosodic, &amp; 2 cross modal)</td>
<td>Receptive: Non emotional prosody discrimination, Emotional prosody discrimination, Name the emotional prosody, Conflicting emotional prosody, Match emotional prosody to an emotional face, Match emotional face to an emotional prosody</td>
<td>Affective– happy, sad, angry, fearful, neutral</td>
<td>164 participants ages 17–85 years, right handed, primarily Caucasian, living in the Southeastern U.S., no psychopathology at the time of testing. Norms are provided for young adults (N = 53, 18–30 years), middle-age adults (N = 42, 31–60 years), older adults (N = 49, 61–70 years) &amp; elderly adults (N = 20, 71–84 years) for each of the 10 subtests (Bowers et al., 1999) &amp; individuals with neurologic disorders (Blonder, Bowers, &amp; Heilman, 1991; Bowers, Blonder, Slomine, &amp; Heilman, 1996).</td>
<td>2 weeks test–retest reliability in young adults (N = 20, 18–30 years) &amp; middle age adults in their early 50s (N = 12) ranged between 0.89 to 0.97.</td>
<td>Time: ~60 min Format: electronic administration; manual scoring Made for: clinical &amp; research use</td>
</tr>
</tbody>
</table>
Table 3 (p. 3 of 3). Features of the assessment tools that evaluate affective prosody only.

<table>
<thead>
<tr>
<th>Test</th>
<th>Age (in years)</th>
<th>Purpose and population</th>
<th>Subtests</th>
<th>Prosody subtests receptive/expressive</th>
<th>Aspects of prosody</th>
<th>Normative sample and inclusion criteria</th>
<th>Reliability/Validity</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS (Pearson, 2009)</td>
<td>16–70</td>
<td>To assess social functioning deficits in adults.</td>
<td>4 subtests</td>
<td>Receptive: Prosody–face matching, Prosody–pair matching</td>
<td>Affective–happy, sad, angry, surprised, fearful, disgusted, neutral, sarcastic</td>
<td>800 participants aged 16–70 years, no history of neurological, psychiatric, developmental, or medical condition affecting cognitive functioning. Demographics matched to U.S. 2005 census data for ethnicity &amp; education level.</td>
<td>Internal consistency: social perception total (α = 0.70–0.84), prosody (α = 0.64-0.79), and pairs (α = 0.78-0.85).</td>
<td>Time: ~30-45 min</td>
</tr>
</tbody>
</table>

Note. MNTAP = Minnesota Tests of Affective Processing, DANVA 2 = Diagnostic Analysis of Nonverbal Accuracy 2, FAB = Florida Affect Battery, ACS = Advanced Clinical Solutions.
Rutherford, 2007; Stojanovik, 2010). The original PEPS (Profiling Elements of Prosodic Systems) test described in Peppé’s (1998) dissertation was designed for clinical use with adults; however, the present version of PEPS–C is specifically designed for use with children. Data on typical adults using PEPS are reported in Peppé, Bryan, Maxim, and Wells (2000), and some clinical results for adults are reported in Peppé, Bryan, Maxim, and Wells (1997). Martinez-Castilla, Sotillo, and Campos (2011) used the Spanish version of the PEPS–C (Peppé et al., 2010) to assess the prosodic abilities of Spanish-speaking adolescents and adults with Williams syndrome. The PPAT and MNTAP have been used to assess prosodic impairments in children with hearing loss and attention deficit hyperactivity disorder (Klieve & Jeanes, 2001; Shapiro, Hughes, August, & Bloomquist, 1993).

The Aprosodia Battery, FAB, and ACS have been used mainly to assess affective prosodic difficulties in adults with conditions such as brain injury and neurologic and psychiatric disorders. Nowicki (2006) reported that the DANVA 2, which was developed for use with children and adults, has been widely used for research purposes in various clinical groups, including hearing loss, ASD, traumatic brain injury, and learning disability.

Crystal (2009) reported that there is inadequate diagnosis in terms of identifying prosodic difficulties in children (mastering prosodic contrasts) and adults (managing the organizational role of prosody in speech production) and in determining the prosodic difficulties caused by access to limited auditory information in children and adults (e.g., due to hearing loss). Our review indicates that tools such as the PEPS–C, DANVA 2, PROP, and PVSP are applicable for wide age ranges (see Table 2) and diverse clinical populations. There is, in our opinion, sufficient information to indicate the intended purpose and target population for these tools but a dearth of published studies exploring their clinical utility. Perhaps the best way forward is to encourage researchers to use these tools to generate evidence for clinical use.

**Domains of Prosody Assessed**

Turk (2009) reported that an ideal prosody assessment tool should involve the assessment of function, phonological representation, surface-level implementation, and perception of prosody (i.e., a comprehensive assessment should be able to evaluate both receptive and expressive skills across different aspects of prosody). An important variation among the assessment tools reviewed here is the attributes of prosody that they evaluate. This review reveals PEPS–C as a comprehensive tool that is useful to assess the perception and production of different aspects of prosody. The receptive component of the PEPS–C includes subtests to assess sentence type (question vs. statement; Turn-End Reception), speaker’s attitude (liking or disliking of food items; Affect Reception), phrase boundaries (the distinction between simple and compound nouns and groupings of adjectives; Chunking Reception), placement of contrastive stress/accent (Contrastive Stress Reception), and auditory discrimination for long (Long item Discrimination) and short (Short item Discrimination) tones. The expressive component includes six subtests analogous to the receptive subtests. The PPAT assesses the perception of phonetic features of prosody such as pitch, duration, and intensity (Apart from Linguistic Context) and grammatical (Grammatical Class and Compound and Abutting Words), emotional (Tone and Affect), and pragmatic (Stress) aspects of prosody. Thus, the PPAT is useful for the comprehensive assessment of receptive prosodic skills in children. However, the PPAT does not assess expressive prosody. Only the MNTAP, FAB, ACS, and DANVA 2 assess the perception of vocal emotions conveyed using prosodic cues. The Aprosodia Battery includes receptive and expressive subtests but assesses only affective prosody. The PROP and PVSP include only expressive prosody subtests to evaluate features such as pitch, phrasing, stress, and loudness.

The Aprosodia Battery and the FAB focus on evaluating affective prosodic deficits in neurologic or psychiatric patients. However, difficulties in affective and other aspects of prosody are observed in other conditions such as hearing loss (Hopyan-Misakyan, Gordon, Dennis, & Papsin, 2009; Most & Peled, 2007). Individuals with sensorineural hearing loss have difficulties perceiving subtle changes in pitch, loudness, and duration (Moore, 1987; Moore & Carlyon, 2005), which are major acoustic cues for the perception of prosody in English. Inaccurate perception of these acoustic cues by individuals with hearing loss are manifested as difficulties in perceiving different aspects of prosody such as differentiating question from statement, word stress, distinguishing word/phrase boundaries, and vocal emotion recognition (Hopyan-Misakyan et al., 2009; Meister et al., 2009; Most & Peled, 2007). The PPAT uses a phonetic and phonological perspective to describe receptive prosodic skills in children using cochlear implants. This is appropriate for individuals with hearing loss because it considers both acoustic factors and linguistic functions of prosody. Unfortunately, prospective trials to evaluate the clinical utility of this tool are missing. Thus, there is a lack of comprehensive, valid, and reliable assessment tools to assess different domains of prosody in individuals with hearing loss.
Normative data. None of the tools described in this review is standardized. However, norms are provided by test developers for the PEPS–C, DANVA 2, ACS, FAB, and PVSP. Norms for the PPAT, Aprosodia Battery, and MNTAP were obtained from studies that used these measures to compare the performance of disordered populations and control groups (see Tables 2 and 3). Normative data for these tests were obtained for sample sizes ranging from 16 to 800.

Diehl and Paul (2009) reported that compared to the PROPP, PVSP, and PEPS–C, language assessment measures such as the CELF–IV and PPVT were normed on larger normative samples with stratified norms based on gender, race, geographic location, and other factors (Dunn & Dunn, 2007; Strauss, Sherman, & Spreen, 2006). Stratified normative data are not available for the prosody assessment tools identified by this review. Clinicians need to be cautious in implementing the normative scores provided by the test developers to their target clinical population. For example, the PEPS–C was normed on a sample of 120 British English-speaking children ages 5–14 years, and the DANVA 2 Child Paralanguage subtest was normed on North Americans ages 8–10 years. These norms may not be appropriate for use with children who belong to different ethnic groups as there are linguistic and cultural prosody differences even among speakers of English (Coggshall, 2008).

Validity. The PVSP test developers reported good instrumental and perceptual criterion validity (Shriberg et al., 1992). Instrumental procedures were used to estimate the criterion validity of more than 300 audiotaped exemplars that were selected to teach the coding procedures. Where the criterion validity of these perceptual coding decisions could not be determined by instrumental means, comparisons with the perceptual decisions of a panel of expert listeners were used. Very few studies have used the MNTAP, FAB, DANVA 2, PEPS–C, ACS, Aprosodia Battery, and PPAT tools to discriminate between typical and atypical populations (discriminant validity; Bowers et al., 1999; Klieve & Jeanes, 2001; Ross et al., 1997; Shapiro et al., 1993), which suggests a need for further validation.

One aspect of validity, face validity, can be addressed by determining the relevance of test items to real-life communication. For example, the Affect and Turn-End Reception subtests of the PEPS–C use single-word test items (names of food items) rather than a sentence context. A positive feature of the ACS, DANVA 2, and FAB is that they use sentence-level stimuli, which are more naturalistic than word-level stimuli. While assessing prosodic skills, it would be appropriate to balance the advantages of psychometric robustness (using normed tests) against the advantages of ecological validity (as in careful analysis and profiling of naturalistic conversational data).

Reliability. Data on the reliability of the nine assessment tools reviewed in this study are shown in Tables 2 and 3. Reliability has generally been assessed using internal consistency measures or measures that are based on the correlations between different items on the same test. For the DANVA 2 and ACS, internal consistency was demonstrated by computing Cronbach’s alpha. Further empirical confirmation of the tool beyond its initial construction by the original developers was undertaken for DANVA 2 (Nowicki, 2006). Internal consistency was measured using the coefficient of variation for the Aprosodia Battery. Test–retest reliability data are provided for the FAB and PEPS–C. Good intrarater reliability (i.e., degree of agreement among raters) and interrater reliability (i.e., consistency of a measure when administered by different examiners) were reported for the expressive subtests of the PEPS–C by Wells, Peppe, and Goulandris (2004). This tool has been used by an increasing number of researchers (Foley et al., 2011; Martínez-Castilla & Peppe, 2008; Stojanovik, 2010).

There is considerable difficulty in directly comparing the different tools described in this review. For example, the sensitivity and specificity of each test varies according to the population studied and the cut-off scores that are considered abnormal (see Tables 2 and 3). Overall, this review supports the recommendation by Diehl and Paul (2009) that there is a need for a prosody assessment tool that (a) is standardized relative to a large representative normative sample and (b) provides empirical data on reliability, validity, and other psychometric properties.

Feasibility

As noted by Crystal (2009), prosody is often neglected in terms of assessment, and “it is difficult to think of another medical area where a set of potentially relevant symptoms would be treated with such unconcern” (p. 257). Green and Tobin (2009) reported that a phonetic (surface-level features) and phonological (functions) analysis of prosody is useful in both typical and atypical speech. Given the clinical relevance of prosody, clinicians should take the time to investigate prosodic skills routinely in their client groups despite time constraints. However, practical considerations are important in a clinical setting, affecting both clients and the professionals involved in the assessment process, especially as time is necessarily limited. Clinicians working in the field might need to consider having access to equipment such as
a laptop, loudspeaker, sound-level meter, and digital voice recorder that would be required to administer some of the tools (e.g., PROP, PVSP, PEPS–C, DANVA 2). Automatic and computerized tests such as the PEPS–C and DANVA 2 are simple, the user manuals provide clear instructions, and a minimum amount of training is required for clinicians. In contrast, the MNTAP, PROP, and PVSP require manual scoring and prosody transcription. Other factors relevant to the clinical utility of the assessment tools include the time taken to administer the test, ease of scoring (manual or computerized), appropriateness of the test stimuli for use with children (color photographs or black and white photographs), test format (computerized or paper pencil test), and age range. A clinician should consider these factors before selecting a prosody assessment tool.

**Administration time.** The time taken to administer each test varies depending on the number of subtests involved and the population being assessed (typical or disordered population). The PEPS–C is a comprehensive test that takes approximately 45–60 min to administer both the receptive and expressive subtests. Wells and Local (2009) described the PEPS–C as not time consuming, whereas Diehl and Paul (2009) considered it to be very long for a clinical measure. Peppé (2009) reported that the PEPS–C is short compared with the process of conversation analysis, and long considering that it tests only prosody. The DANVA 2, MNTAP, FAB, and ACS take relatively less time (approximately 20–30 min) compared to the PEPS–C but only assess the perception of affective prosody. The PPAT and Aprosodia Battery take approximately 1 hr to administer. Longer test duration would not be appropriate for young children and some clinical populations such as attention deficit hyperactivity disorder or ASD. Tests that are in paper–pencil format and require manual scoring (MNTAP, FAB, PPAT) would take longer for the clinician than automatic and computerized tests (PEPS–C receptive subtests, DANVA 2). The time taken to transcribe prosodic elements (PROP, PVSP, expressive subtests of Aprosodia battery, and PEPS–C) would depend on the expertise of the clinician.

**Appropriateness.** The appropriateness of the test stimuli should be considered, particularly when assessing children and atypical populations. Color photographs as in the PEPS–C and DANVA 2 would appear more realistic and have higher ecological validity (Diehl & Paul, 2009) than black and white photographs (MNTAP). The color of test stimuli can have a positive effect on performance levels in children (Jeanes et al., 1997). The number of response items can affect the chance performance level and the cognitive demands of the task. The PEPS–C receptive subtests use a simple two-alternative forced-choice format (50% chance performance level). The ACS, DANVA 2, and FAB tools have a minimum of four response options (25% or lower chance performance level), and hence may not be feasible for very young children and some clinical populations.

The version of English that is used to record the test stimuli can differ from that of the target population; hence, locally developed norms may be required for prosody assessment tools using audio-recorded material. The DANVA 2 test stimuli were recorded by native English speakers from the United States, so its suitability for assessing prosodic skills in speakers of other versions of English such as Australian or New Zealand English needs further evaluation. A positive feature of the PEPS–C tool is that the recordings are available in four different versions of English, including British, Australian, North American, and Scottish English. Clinicians need to consider the cross-dialect prosodic variations while assessing speakers of Afro-Caribbean, Singaporean, or Indian versions of English.

**Sensitive to development.** Diehl and Paul (2009) reported that there is a need for a prosody assessment tool that is developmentally sensitive and can be used with different age groups. Of the tools reviewed here, only the DANVA 2 has different forms for different ages. Tools without different age versions can have ceiling effects for older children and adults and floor effects for younger children (e.g., Wells & Peppé, 2003). Also, different subtests may not be equally difficult for the age group tested, making it difficult for the clinician to determine relative strengths and weaknesses across areas for the purpose of intervention unless good normative data are available (Diehl & Paul, 2009). For example, the PVSP can be used for a wide age range but does not indicate what percentage of correct prosody would be appropriate for various age groups. Few studies have reported age-related developmental changes on prosodic skills in children using the PEPS–C (Foley et al., 2011; Gibbon & Smyth, 2013; Wells et al., 2004). A summary of the results and discussion section is provided in Table 4.

**Conclusion**

SLPs frequently encounter prosodic impairments in persons with various communication disorders; hence, knowledge regarding assessment options for persons with prosodic impairments is important. The aims of this review were to identify the tools that are available to assess prosodic skills in children and adults and to evaluate the clinical utility of each. In recent years, methodological paradigms such as acoustic
analysis of speech productions, direct measurement of articulatory movements, judgments and reaction times obtained during identification and discrimination tasks, measurements of brain activity, and patterns of attention in babies have been used in prosody research (Prieto, 2012). However, these techniques are time consuming and are not feasible in a clinical environment. This article described the nine behavioral assessment tools that are available to aid clinicians who wish to examine prosodic skills in typical and disordered populations. The relatively small number of tools available to evaluate prosody compared to other aspects of language indicates that, although prosody is a topic that is clinically relevant, it is often overlooked in terms of formal assessment. Clearly, there is no widespread recognition of the need for prosody assessment, as evidenced by the small number of assessment tools available.

Consistent with Diehl and Paul (2009), our literature review identified very few assessment tools available to evaluate prosody compared to the large number of standardized assessment tools that are available for other aspects of language like syntax, vocabulary, and phonology. Many existing prosody tools are narrow in scope or have not been robustly validated. This is an important gap that warrants future research. Most tools have been carefully constructed but lack generalizability across prosodic disorders, and five out of the nine tools identified focus on only one particular aspect of prosody. If a clinician is working with client groups with neurologic impairments and psychiatric disorders in which affective prosody is the main focus, then Table 3 will be useful.

When selecting a tool, clinicians should consider the target population, time required to administer the tool, ease of administration and format of the tool, and access to equipment, as well as normative and reliability data. The various assessment tools reported in this review have explained prosody using different perspectives; hence, clinicians should have a specific idea about which aspect of prosody they want to assess and on which dialects of English it can be used safely.

The assessment of prosody is currently constrained by a lack of normative data. Three of the identified tools were devoid of norms. Only four tools are available that focus on two or more aspects of prosody. Among these, the PROP and PVSP require a high level of expertise to transcribe prosodic elements and are time consuming. The PPAT was originally developed for use with children using cochlear implants and therefore focuses only on assessing receptive prosody skills. The PPAT may work well with children with hearing loss, but limited empirical data are available. The PEPS–C may be a good choice in clinical practice as it covers a wide age range, is easy to administer and score, provides a good user guide and manual, and has been used with a number of different clinical populations. However, its application to adults needs further investigation.

A lack of knowledge of developmental norms for different domains of prosody makes it difficult to derive standardized scores for these tools in children. Some reasons why prosody assessment is neglected by clinicians may be a lack of training or awareness of existing assessments, time constraints, lack of normative data for comparison with atypical populations, lack of evidence-based studies regarding the intervention of prosodic difficulties, and lack of culturally appropriate and developmentally sensitive measures. The remarkable lack of published studies on intervention for prosodic deficits might be due to the lack of appropriate assessment tools. This review

<table>
<thead>
<tr>
<th>Tools</th>
<th>Target population</th>
<th>Subtests involved</th>
<th>Psychometric data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Children</td>
<td>Adults</td>
<td>Receptive</td>
</tr>
<tr>
<td>PROP</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>PVSP</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>PEPS–C</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>PPAT</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>MNTAP</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>DANVA 2</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Aprosodia battery</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FAB</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>ACS</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Note. * indicates that norms for control groups were derived from previous studies; # indicates that few studies have used these tools to discriminate between typical and atypical populations.
highlights the needs for prosody assessment tools that are sensitive to developmental changes in children and that comprehensively and reliably assess relevant aspects of prosody in children and adults.

It is worth mentioning here that previous research explaining the articulatory movements and acoustic features associated with various prosodic contrasts (e.g., Grigos & Patel, 2007, 2010; Patel & Grigos, 2006; Snow, 1994, 1998) are not tied to any of the assessment tools described in this review, and a good way forward is to use these theoretical data to develop future tools. Future research should gather empirical data on the acquisition of prosody using the available tools and further explore the establishment of standardized diagnostic measures to evaluate prosody. Validity and reliability have been addressed to varying degrees—thoroughly in some tools and not at all in others.

The effectiveness of these tools in highlighting specific aspects of prosody warranting clinical intervention, and then guiding the intervention approach, has received little attention. Therefore, when considering these tools for the clinical assessment of prosody, caution is required to ensure that the time investment is warranted in terms of improved clinical outcomes for clients with communication difficulties. Given the considerable evidence for the importance of prosody in everyday communication, this is an important area for future work.

**ACKNOWLEDGMENT**

This study was supported by the Oticon Foundation Denmark.

**REFERENCES**


Kalathottukaren et al.: Prosody Assessment Tools 151


Kalathottukaren et al.: Prosody Assessment Tools 153


Contact author: Rose Thomas Kalathottukaren, Building 721.320, Tāmaki Innovation Campus, Discipline of Speech Science, School of Psychology, The University of Auckland, 261 Morrin Road, Private Bag 92019, Auckland, New Zealand. E-mail: rkal180@aucklanduni.ac.nz