Single word and sentence intelligibility in children with cochlear implants

FADWA A. KHWAILEH¹, & PETER FLIPSEN, JR.²

¹University of Tennessee-Health Science Center, Knoxville, TN, USA, and ²Idaho State University, Pocatello, ID, USA

(Received 30 November 2009; Accepted 26 April 2010)

Abstract
This study examined the intelligibility of speech produced by 17 children (aged 4–11 years) with cochlear implants. Stimulus items included sentences from the Beginners’ Intelligibility Test (BIT) and words from the Children Speech Intelligibility Measure (CSIM). Naïve listeners responded by writing sentences heard or with two types of responses to the word recordings; open transcription (CSIM-T) and closed set multiple choice (CSIM-MC). Percentage of items understood, averaged across three naïve judges, were compared across the three measures. Additionally, scores were examined for any relationships with chronological age, age of implantation, and amount of implant experience. Strong positive correlations were observed among all three intelligibility tasks. Scores on all three tasks were found to be significantly different from each other. A significant correlation was obtained between intelligibility and amount of implant experience, but not with chronological age or age of implantation. Results suggest that judging the intelligibility of speech produced by children with cochlear implants using both single word and sentence levels would provide a better overall estimate of their intelligibility. The results also emphasize the vital role of auditory input in the development of intelligible speech.

Keywords: cochlear implants, intelligibility, children

Introduction
Making oneself understood is the ultimate goal of human communication, and failure to establish fully intelligible speech may lead to a significant communication handicap. The measurement of intelligibility is therefore crucial to much of speech pathology practice. In particular, measures of speech intelligibility may be necessary for (a) determining if intervention is needed, (b) setting intervention goals, and (c) evaluating the effectiveness of intervention (Ansel, McNeil, Hunker, and Bless, 1983; Bernthal and Bankson, 1998; Monsen, 1983; Schmidt, 1984). Given the importance of such measures in making clinical decisions for children, they need to be accurate, reliable, and valid (Gordon-Brannan and Hodson, 2000).

Poor speech intelligibility has long been an issue for individuals with hearing impairment. Gold (1980), for example, reviewed several studies of spontaneous speech in this population and noted
‘... only about 20% of the speech output of the deaf is understood by inexperienced listeners’ (p. 397). Problems with speech production for this population include poor accuracy of speech sounds (Levitt and Stromberg, 1983), developmental and non-developmental error patterns (Parker, 2005), inappropriate voice pitch (Angelocci, Kopp, and Holbrook, 1964; Boone, 1966), problems with word duration and pausing (e.g., Parkhurst and Levitt, 1978), and difficulty with intonation (e.g., Allen and Arndorfer, 2000). Smith (1975) and Maassen and Povel (1985) noted that, while both segmental and suprasegmental errors significantly affect intelligibility, segmental errors appear to account for a greater portion of the variance.

Cochlear implants are proving successful for remediating severe and profound hearing loss in children. These devices allow for improvements in the development of speech and language skills in this population (e.g., Blamey, Barry, and Jacq, 2001; Geers, 2004; Geers, Tobey, Moog, and Brenner, 2008; Nicholas and Geers, 2008; Svirsky, Robbins, Kirk, Pisoni, and Miyamoto, 2000; Tobey, Geers, Brenner, Altuna, and Gabbert, 2003). Relative to the current study, a recent review by Flipsen (2008a) indicated that cochlear implants appear to result in intelligibility of the speech that is generally superior to that reported by Gold (1980). An analysis by Flipsen (2009) concluded that, similar to both Smith (1975) and Maassen and Povel (1985), segmental accuracy appears to be the most important remaining factor. Consonant accuracy in particular was found to be the single biggest factor affecting intelligibility in these individuals, with five other variables (accuracy of early consonants, pitch, laryngeal quality, backing errors in word-initial position, and glottal stop substitutions in medial position) together accounting for only ~ 3.1% of the variance.

Despite the improved intelligibility outcomes overall in children with cochlear implants, considerable variability remains across individuals. Classically the variability observed in those with hearing loss has been attributed to such factors as age of onset of hearing loss, degree of hearing loss, amount of amplification experience (i.e., hearing age), and type of intervention. Although age of onset and degree of loss remain important issues, most recent studies of children with cochlear implants have focused on those with congenital losses (i.e., those identified very early, often at birth) which are severe-to-profound in extent (i.e., those meeting usual eligibility requirements). The remaining emphasis has been on three age-related variables, age of implantation, amount of implant experience, and chronological age. Chin, Tsai, and Gao (2003), for example, reported significant correlations between sentence-level intelligibility and both chronological age (r = 0.71), and amount of implant experience (r = 0.65). Age of implantation did not appear to have been considered in that study. Flipsen and Colvard (2006) also reported significant correlations between conversational intelligibility and chronological age (r = 0.64) and with amount of implant experience (r = 0.49). Both Miyamoto, Svirsky, Iler Kirk, Robbins, Todd, and Riley (1997) and Osberger, Robbins, Todd, Riley, and Miyamoto (1994) reported increases in sentence level intelligibility with length of device use, although no statistical correlations were reported. Relative specifically to age of implantation, most studies report superior intelligibility outcomes in those who receive their implants earlier rather than later (Löhle, Frischmuth, Molm, Becker, Flamm, Laszig, et al., 1999; Loundon, Busquet, Roger, Moatti, and Garabedian, 2000; Miyamoto, Iler Kirk, Svirsky, and Sehgal, 1999; Tye-Murray, Spencer, and Woodworth, 1995). These studies did not specifically report correlations between intelligibility and age of implantation. Interestingly, Huttunen (2008) reported no significant correlation between single word intelligibility and age of implantation.

Measuring intelligibility

A paramount concern of researchers and clinicians—especially those who work with young children—has been the development of a valid means of measuring speech intelligibility.
Three common approaches are open-set word identification, closed-set word identification, and rating scales with equal intervals procedures (Gordon-Brannan, 1994). The first two methods yield intelligibility percentages. Open-set methods are based on listener identification through orthographic or phonetic transcription of the words spoken (single words, sentences, or continuous speech) by the listeners. Closed-set methods involve the listener selecting words from a pool of choices.

The source of difficulty in selecting the appropriate approach to use comes from the procedural and interpretative complications that accompany intelligibility evaluation (Kent, Miolo, and Bloedel, 1994). Because determining actual percentages of words understood through word identification tasks can be time-consuming (Bacon, 1995), they are not done routinely by clinicians. Impressionistic estimates of intelligibility using rating scales suffer from significant validity and reliability concerns (Schiavetti, 1992). Word identification tasks also offer the additional advantage of providing a starting point for identifying possible sources of intelligibility deficits. Thus, they have the potential to go beyond merely indexing severity (Schiavetti, 1992).

Measuring intelligibility in children with cochlear implants

As with other populations, studies of children with cochlear implants have varied relative to how intelligibility is measured. For example, speakers in these studies have variously produced single words (Chin, Finnegan, and Chung, 2001; Löhle et al., 1999; Mondain, Sillon, Vieu, Lanvin, Reuillard-Artieres, Tobey, et al., 1997; Vieu, Mondain, Blanchard, Sillon, Reuillard-Artieres, Tobey, et al., 1998), sentences (Chin et al., 2001; 2003; Dawson, Blamey, Dettman, Rowland, Barker, Tobey, et al., 1995; Miyamoto, Iler Kirk, Robbins, Todd, and Riley, 1996; Miyamoto et al., 1997; 1999; Osberger, Robbins, Todd, and Riley, 1994; Tye-Murray et al., 1995), narratives (Tye-Murray et al., 1995), or conversational speech (Allen, Nikolopoulos, and O’Donoghue, 1998; Calmels, Saliba, Wonna, Cochard, Filliaux, Deguine, et al., 2004; Flipsen and Colvard, 2006; Inscoe, 1999; Löhle et al. 1999; Loundon et al., 2000). Few studies have included across-task comparisons of intelligibility; however. Chin et al. (2001) compared sentence level productions from the Beginners’ Intelligibility Test (BIT; Osberger, Robbins, Todd, and Riley, 1994) with closed-set minimal pairs from the Minimal Pair Test (MP2; Robbins, Renshaw, Miyamoto, Osberger, and Pope, 1988) in paediatric users of cochlear implants. Results indicated that sentence intelligibility was significantly correlated with single word intelligibility ($r = 0.77$). One limitation of the Chin et al. study was that the tasks differed on two dimensions simultaneously. As such it was not possible to separate the influence of the judgement task (transcription vs closed set choice) from the influence of the linguistic level of the material (single words vs sentences).

An additional challenge for measuring intelligibility with individuals who have hearing impairments occurs when using structured, imitative tasks (as in the current study). Such tasks offer the distinct advantage of providing known targets and thus a controlled basis for calculating the percentage of speech understood by unfamiliar naïve listeners. However, the use of imitative tasks for measuring intelligibility (i.e., a measure of output) at the single word level assumes that the stimuli have been at least appropriately perceived by the auditory system. At the phrase or sentence level and above it also assumes that the stimuli have not been modified in any way because of limitations of the cognitive-linguistic system. At least two approaches are possible to deal with these assumptions. The first would be to score performance relative to actual output rather than the original target. For example, if the imitative target was ‘She is taking a bath’, a speaker with a hearing impairment might omit the
auxiliary, present progressive-ing and the indefinite determiner and produce ‘She take bath’. Such a production might indicate a perceptual problem, a language impairment, or some combination of both. Scoring relative to actual output would prevent penalizing the speaker for such problems but only evaluate output abilities. A listener transcription of ‘she take bath’ then yields a score of 3/3 (100%) because the listener successfully understood what the speaker actually produced. However, that approach fails to consider that intelligibility above the single word level is a product of many factors including language formulation skills. While the listener may have comprehended the core content of the output ‘she take bath’, and even grasped the speaker’s intention for that particular message, such a production in a larger context such as conversation or narrative may actually distract the listener enough to interfere with understanding other parts of the speaker’s output. The alternative to such a modified scoring is to compare the listener transcriptions to the original target but keep in mind the fact that the obtained intelligibility scores actually reflect a combined measure of input, processing, and output abilities.

The current study

The purpose of this retrospective study was to compare intelligibility scores of children with cochlear implants across two linguistic levels (single words and sentences); single word productions were scored using both open set and closed set procedures. This would allow for direct comparison of the two scoring methods as well as allow contrasting performance across linguistic levels. In addition, the study sought to add to our understanding of the pattern of emergence of intelligible speech by examining all three sets of intelligibility values relative to the influence of chronological age, age of implantation, and amount of implantation experience. Naïve listeners were used as judges of intelligibility to avoid possible familiarity bias.

Methods

Participants

Participants were 17 children (four males; 13 females) recruited from Child Hearing Services at the University of Tennessee (see Table I). Based largely on a review of available records, all were found to be pre-lingually deaf with bilateral, severe, or profound sensorineural hearing loss (i.e., >70 dB HL). Pre-implant evaluations had been conducted at other facilities, and thus most files did not include pre-implant audiograms or any other specific information about how hearing loss categorization had been determined. No other handicapping issues were reported. Precise information on age of identification was not available in many cases. Age at time of testing ranged from 4;8–11;1 (mean = 8;0; median = 7;9). Age of implantation ranged from 1;2–8;4 (mean = 3;9; median = 3;0). Amount of implant experience ranged from 1;0–7;10 (mean = 4;3; median = 4;8). Information on specific devices and/or speech processors was not collected at the time of the recordings and could not be ascertained for most participants later (several were no longer being seen at the test facility). All had used hearing aids for a minimum of 3 months (and some for up to several years) prior to implantation. It should be noted that participants 4 and 5 were identical twins. It should also be noted that the uneven gender distribution in the current sample is likely not typical of this population, as most studies using larger samples (e.g., Geers, 2004) have included relatively equal numbers of males and females.
As indicated in Table I, the children were implanted with a mix of implant devices. All but one had unilateral implants. Participant 8 was a bilateral user, but had used the second implant for only 5 months at the time of testing. The participant was therefore retained in the study because the amount of time using the second implant was judged to likely have had minimal impact on speech production. Age of first implantation (shown in Table I) was used in the calculations herein. Participant 11 had experienced a partial device failure with a re-implantation 12 months prior to testing. Children with left and right ear implants were included; direct comparisons of findings for different ear of implantation are presented in Flipsen (2008b). In order to determine whether ear of implantation would confound the current results, a series of Mann-Whitney tests were carried out; no significant difference was found between the participants with right ear implants compared to those with left ear implants in terms of chronological age, age of implantation, or amount of implant experience (all \( p > 0.05 \)).

Data collection

The Children’s Speech Intelligibility Measure (CSIM; Wilcox and Morris, 1999) was administered to evaluate single word intelligibility. The CSIM contains 50 word lists each containing 12 similar sounding words. Prior to administering the CSIM, the examiner randomly chose a word from each of the 50 lists. Thus, a different unique set of words was chosen for each child. Each child imitated the 50 words spoken by the examiner.

The Beginners’ Intelligibility Test (BIT; Osberger, Robbins, Todd, and Riley, 1994) was administered to evaluate sentence level intelligibility. The BIT was chosen because it contains four lists of 10 sentences, each containing words familiar to children; the sentences range in length from two-to-six words (mean = 3.8; median = 3.5) and from three-to-eight syllables.
(mean = 4.5; median = 4.0), lengths which would be manageable for many young children. Each of the four lists of 10 sentences contains between 37–40 total words (mean = 38.3; median = 38). Pictures are used to make the context more meaningful for the child. Each child was randomly assigned one of the four BIT sentence sets. The examiner would recite each sentence while showing the picture to the child who would then repeat the sentence.

Both the CSIM and the BIT were administered in fall 2005 as part of a larger protocol for each child during one session by a single, trained, graduate student examiner who tested all of the children. The duration of each testing session did not exceed 60 minutes. All sessions were conducted in a single wall, sound treated booth.

Productions of both the CSIM and BIT were recorded on digital audiotape using an external microphone attached to the hand of a puppet sitting on the table approximately two feet from the participant. Although this distance is somewhat greater than used in similar studies, it was deemed necessary to keep some of the youngest participants from handling the microphone and thus introducing extraneous noise. The distance was not considered a problem given that (1) recordings were being conducted in the highly-controlled environment of a sound booth, and (2) the recordings were later to be amplitude-normalized prior to listener judgements. At the beginning of each recording session recording levels were established at optimum (i.e., highest possible level without the needles hitting maximum) from the child’s initial output using the VU meters of the tape recorder.

Following test administration, the recorded productions were transferred to a laptop computer and each of the 50 single-words from the CSIM and the 10 sentences from the BIT were isolated into independent digital files using Computerized Speech Lab (CSL 2001, Model 4400, Kay Elemetrics, Corp. Lincoln Park, NJ). Once isolated, all files were RMS amplitude normalized using the software program Cool Edit Pro (1999, Version 1.2, Syntrillium Software, Corp. San Jose, CA).

Judges

A group of University of Tennessee undergraduate students (n = 102) participated to receive extra credit in a psychology class with 51 judges for the CSIM and a different set of 51 judges for the BIT. The judges’ ages ranged between 18–23 years. All were native speakers of English who reported having no or minimal experience with the speech of the hearing impaired.

For the CSIM, judges orthographically transcribed the 50 items from one speaker and then immediately thereafter judged a second speaker using a multiple-choice format (they were given blank copies of the 50 sets of 12 and asked to circle the word heard). Prior to listening sessions, the order of the speakers was arranged so that listeners judged two children with different sets of words. For the BIT the judges transcribed the 10 sentences from each of three different speakers (always three different word lists). Each item on both the CSIM and BIT was presented twice in succession. In the transcription tasks, subjects were asked to write what they thought they heard.

Naïve judges were used in the current study to avoid potential for bias by experienced listeners. Judges listened to no more than three different speakers each in order to avoid the possibility of becoming too familiar with the speech of this population.

Listening sessions were conducted in the same sound booth where recordings had been made. Listening sessions were made individually or in groups of two-to-three. A total of three judges were used for each of the samples.
Data analysis

Percentage understood values were calculated for each listener for each task and then averaged across the three listeners who heard the sample. In the case of the CSIM, scores for both transcription (CSIM-T) and multiple choice (CSIM-MC) were calculated separately.

In the case of BIT, the number of words correctly identified was calculated based on a match between the target sentence and the listener judges’ responses. For example, if the target sentence is *She is cooking dinner*, and the child says ‘She is cook dinner’, omitting the progressive suffix -ing, a transcribed response of *She is cook dinner* scores 3 of 4. Data on the language comprehension abilities of the participants were not complete. Because we could not assume age-appropriate language comprehension, we opted to score production relative to the original stimuli.

The process of scoring yielded three intelligibility scores for each speaker (CSIM-T; CSIM-MC; BIT). The three scores were compared with each other using repeated measures one-way ANOVA with a Bonferroni-corrected $p < 0.017$ (.05/3) used to test for statistical significance. The three intelligibility scores were also each correlated with three age variables (chronological age, age of implantation, amount of implant experience) again using a Bonferroni-corrected $p < 0.017$ (.05/3) to evaluate statistical significance.

Reliability

Inter-judge reliability was calculated two ways. First, a fairly conventional approach was taken in which Pearson correlations were calculated between the highest and lowest scores across the three judges for each task. The obtained correlations were 0.98, 0.98, and 0.98 across the three tasks (CSIM-T, CSIM-MC, BIT), respectively. This suggested a relatively high level of inter-judge agreement. However, correlations may tend to mask differences for individual scoring events. Therefore, a second approach was taken which involved examining the range of score differences across judges. This was evaluated by comparing the highest and lowest scores of the three judges for each intelligibility task separately for each speaker. On the CSIM-T, differences between the highest and lowest score ranged from 2–16% ($M = 7.5\%; SD = 3.6\%$). On the CSIM-MC, differences ranged from 2–16% ($M = 8.7\%; SD = 4.5\%$). On the BIT, differences ranged from 2.7–18.9% ($M = 10.6\%; SD = 6.1\%$). The degree of variability across judges supports the idea of using multiple judges in the clinical application of these tests.

Results

The average intelligibility scores for each of the participants on each of the tasks are shown in Figure 1. As a group, intelligibility on the CSIM-T ranged from 1.3–77.3% ($M = 32.8\%;$ Median = 32.7; SD = 1.3). On the CSIM-MC, intelligibility ranged from 33.3–89.3% ($M = 66.7\%;$ median = 72.7; SD = 19.1). On the BIT, intelligibility ranged from 9–96.5% ($M = 57.4\%;$ median = 61.7; SD = 27).

A repeated measures one-way ANOVA revealed that there were significant differences between the three intelligibility scores ($F_{(1,040)} = 53.266, p < 0.017$). LDS comparisons revealed that all three means were significantly different from each other (all $ps > 0.017$).

Scores on the CSIM-T were always lower than scores on the CSIM-MC. The two sets of scores were found to be significantly correlated ($r = 0.85; p < 0.017$). Scores on the CSIM-T
were also always lower than scores on the BIT. Once again, the two sets of scores were significantly correlated ($r = 0.78; p < 0.017$). Thirteen out of 17 scores on the CSIM-MC were higher than scores on the BIT. The two sets of scores were also significantly correlated ($r = 0.89; p < 0.017$).

The relationships among the three intelligibility tasks and the age variables are shown in Table II. The only significant correlations ($p < 0.017$) were obtained between amount of implant experience and all three variables, although the correlations tended to be moderate. The significant correlations accounted for 34–45% of the variance. Intelligibility at both the single word and sentence levels was associated with amount of implant experience, but not with age of implantation or chronological age.

Table II. Pearson correlations between intelligibility measures and age variables.

<table>
<thead>
<tr>
<th></th>
<th>Chronological age</th>
<th>Implantation age</th>
<th>Implant experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSIM-T</td>
<td>.28</td>
<td>.34</td>
<td>.58*</td>
</tr>
<tr>
<td>CSIM-MC</td>
<td>.32</td>
<td>.41</td>
<td>.67*</td>
</tr>
<tr>
<td>BIT</td>
<td>.42</td>
<td>.30</td>
<td>.64*</td>
</tr>
</tbody>
</table>

CSIM-T = Children’s Speech Intelligibility Measure (Wilcox and Morris, 1999), scored by transcription; CSIM-MC = Children’s Speech Intelligibility Measure, scored as multiple-choice task; BIT = Beginner’s Intelligibility Test (Osberger, Robbins, Todd, and Riley, 1994).

* $p < .017$. 

Figure 1. Average intelligibility scores (percentage understood) for the 17 participants across the three intelligibility tasks.
Discussion

The present study investigated speech intelligibility of a group of 17 children fitted with cochlear implants across two linguistic levels: single words and sentences using three tasks. Performance among the three tasks was significantly positively correlated, although all of the scores differed from each other. The high positive correlations among CSIM-MC, CSIM-T, and BIT suggest that each of these three intelligibility tasks would provide similar rankings of the children’s intelligibility. Thus, there is high concurrent validity support among the tasks. This finding also provides support for the construct validity of the two tests (i.e., both the CSIM and the BIT appear to offer valid estimates of the intelligibility of children with cochlear implants).

The magnitude and direction of the correlation obtained between BIT and CSIM-MC ($r = 0.89$) is similar to the findings of Chin et al. (2001), who reported significant correlations between BIT scores and their Minimal Pair Production ($r = 0.77$) task.

Scores on the CSIM-T were significantly lower (and lower in all cases) than scores on the CSIM-MC or the BIT. Scores on the CSIM-MC were significantly higher than scores on the BIT. Even though the sentence level nature of the BIT provides contextual (i.e., meaning) cues to assist the listener, it likely makes the listening task harder by requiring the listener to deal with more information. On the other hand, the CSIM-MC offers no such meaning cues, but the multiple choice task narrows the field of choices for the listener. Thus, the relative advantages and disadvantages of each may explain the higher scores on CSIM-MC than those on the BIT. However, BIT scores were always higher than CSIM-T scores. This highlights the fact that the contextual cues provided by the BIT clearly offer an advantage to the listener. The significant difference between CSIM-T and CSIM-MC scores again indicates that the narrow set of choices the closed-set task offers to the listener likely contributes to its higher scores, even though both tasks lack contextual cues. These findings are consistent with Yorkston and Beukelman (1978), who reported lowest scores on transcription tasks and highest scores on multiple-choice tasks.

The developers of the CSIM recommend the use of the closed-set (i.e., multiple-choice) format. The fact that the values of CSIM-MC, CSIM-T, and BIT were significantly different from each other supports the notion that all speakers really have a range of intelligibility potentials (Kent et al., 1994). Thus, capturing the range of intelligibility of children with cochlear implants by employing both single word productions and sentence productions would give us a better sense of their intelligibility range. In addition, findings from the current study would support a recommendation that both should involve open-set tasks. If one accepts Kent’s (1993) ‘triangulation’ model, the addition of conversational speech intelligibility assessment would more completely reflect the overall communicative ability for children. Nonetheless, the issue of the best approach of quantifying conversational speech intelligibility is still not fully resolved (Flipsen, 2006). Because it provides a frame of reference on the exact number of words attempted, the addition of sentence intelligibility assessment might be more convenient for the clinician. The inclusion of more than one linguistic level in determining intelligibility (assuming the same scoring approach is used) may assist the clinician in isolating some of the factors that are affecting intelligibility.

Consistent with previous studies (e.g., Chin et al., 2003; Flipsen and Colvard, 2006; Miyamoto et al., 1997; Osberger, Robbins, Todd, Riley, and Miyamoto, 1994; Tye-Murray et al., 1995), amount of implant experience was significantly correlated with both of the intelligibility measures in the current study. This highlights the critical role of auditory input in the development of intelligible speech in this population.
Findings from the current study should of course be interpreted somewhat cautiously. The convenience sample included only 17 participants as well as a somewhat disproportionate number of females. It is also worth noting that the sample of the study was diverse in terms of chronological age, implantation age, and amount of implant experience. One might expect these large ranges to yield higher correlations with the intelligibility scores than more restricted ranges. In contrast, the correlations in Table II were in the moderate range. This suggests that other factors have contributed to the observed score variance beside the age factors. These factors may be physiological, such as the number of surviving spiral ganglion cells, electrode placement and insertion depth, electrical dynamic range, and signal processing strategies; in addition, other educational, psychological, and social factors such as the recipient’s motivation or level of intelligence may contribute to the individual variability of cochlear implant recipients (Loizou, 1998; Waltzman, 2000). Additional study is clearly indicated.

Conclusion

The current study supports the notion that the assessment of the intelligibility of speech produced by children with cochlear implants should include both single word and sentence level measures. As well, open-set (i.e., transcription) tasks would appear to be preferable in both cases. The use of experienced vs naïve listeners remains to be evaluated.

Acknowledgements

Many thanks to the children who participated in the study and their parents. Special thanks to Rhonda Parker for testing the participants and to the anonymous reviewers for their valuable comments. We also thank the students at the University of Tennessee who served as judges for the BIT and CSIM tasks. Finally, thanks to Dr Molly Erickson and Dr Seunghee Ha for their valuable assistance.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References


