Clarification about the Use of Nonsense Syllable Tests to Assess Auriemmo et al (2009)

We agree with Dr. Jerger’s editorial comments in the May 2009 issue that Auriemmo et al.’s study may offer new possibilities for the use of linear frequency transposition (LFT) in some persons with severe high-frequency hearing losses. However, we wish to comment about the stimuli, methods, and proofreading involved with the study. We are not necessarily criticizing the study or its outcomes but are concerned and confused about the authors’ reported use of the nonsense syllable test (NST) stimuli, which they failed to describe adequately in their methods and referred to erroneously throughout the article. On page 294, the authors stated the following. “The efficacy of LFT for soft and conversational speech intelligibility was assessed at 30 and 50 dB HL input levels using nonsense syllables from the CUNY Nonsense Syllable Test (Edgerton and Danhauer, 1979, Form A, Lists 1 through 6). The full recorded, 25-item lists of CVCV [sic] from its original commercial compact disc were administered for each condition.” Also, on page 296, the authors stated, “The first visit was an assessment using the CUNY Nonsense Syllable Test in an aided condition using the child’s own hearing aids at 30 and 50 dB HL.” The issue of exactly which NST they used requires further clarification prior to accepting the validity of their results. It appears from the authors’ very brief description in the methods and their inconsistent referencing that they were not sure which NST they employed.

In the late 1970s, two different nonsense syllable tests were provided in the literature and for commercial use by researchers and clinicians. Both have been supported by considerable standardization and research over the years. The first was provided by Resnick and colleagues and was based on work at the City University of New York, which was published in 1975 and became known as the City University of New York Nonsense Syllable Test, or CUNY NST (Resnick et al, 1975). The second was based on work by Edgerton, Danhauer, and colleagues at Bowling Green State University and the University of California Santa Barbara, which was first published in 1979 and became known as the Edgerton-Danhauer Nonsense Syllable Test, or ED-NST (Edgerton and Danhauer, 1979). Rationales, details, and considerations for use of both tests were provided earlier by Lucks Mendel and Danhauer (1996).

Both the CUNY NST and the ED-NST are available in analog and digital formats and can be used to assess listeners’ perceptions of speech with materials having normal English phonemes and syllabic structure but lacking semantics in order to tax performance with more difficult stimuli than are found in meaningful word or sentence materials. Although both tests use nonsense syllable stimuli, they involve very different constructions and scoring procedures. That is, the CUNY NST uses consonant-vowel (CV) and vowel-consonant (VC) syllables and a closed-set or forced-choice format where listeners select the item heard from a list of four alternatives (i.e., three foils and the target). This format is often thought to be easier to administer and score than is the open-set or free-response method used with the ED-NST, which uses CVCV syllables. In the ED-NST, listeners are allowed to repeat whatever they think they hear, and the examiner transcribes the responses (usually using phonetic transcription), which has been criticized as being too demanding for some clinicians. Users of the CUNY NST simply score items as right or wrong to arrive at a percent correct score for the whole syllable. However, for those motivated to transcribe listeners’ responses, the ED-NST offers them the ability to score by phoneme, consonant, vowel, and/or word (bisyllabic), and to analyze errors accordingly. Although the CUNY NST might be faster to give and score and the ED-NST may require a bit more initiative on the part of the examiner, the ED-NST can provide a wealth of information about listeners’ phoneme perception in a short amount of time (each list takes only about three minutes to administer), making it attractive for studies like that of Auriemmo et al (2009). In fact, we believe that Auriemmo et al (2009) actually used the ED-NST rather than the CUNY NST based on the reference cited and what we could glean from bits of the description of the test throughout the article. Thus, we will consider that to be the case for this commentary.

Unfortunately, Auriemmo and colleagues omitted so many details about their administration of the test to children with hearing loss in soundfield conditions that further explanation is warranted. At least three earlier studies have investigated use of the ED-NST with children (Danhauer et al, 1985; Danhauer et al, 1986; Johnson, 2000). In all three, children were given the Fisher-Logemann Test of Articulation (Fisher and Logemann, 1971) prior to administration of the ED-NST to ensure that any errors found were perceptual rather than articulatory in origin. Although Auriemmo et al (2009) evaluated their subjects’ language performance and stated on page 293 that “children were all
able to repeat nonsense syllables,” that does not account for speech articulation errors that could have affected the results. They failed to provide information about the speech articulation abilities of their participants. The reading task assessing the children’s productions of /s/ and /z/ with and without hearing aids suggested that their participants did in fact have articulation errors that may have interfered with the validity of ED-NST results.

Likewise, previous studies using the ED-NST with children have stressed the importance of reporting exactly how it was scored. Edgerton and Danhauer (1979) advocated that examiners use both visual and auditory cues when administering the ED-NST, rather than simply monitoring listeners’ responses via headphones or loudspeakers. Johnson (2000) even found it necessary for her examiners to use both auditory and visual cues made available on a TV monitor via a closed-circuit video recording system for ED-NST administration, particularly when scoring the responses of six- to seven-year-old children with normal hearing. She reported that these children frequently mumbled or provided responses that were unintelligible and had syllabic structures different from that of the test stimuli (e.g., CVCVC instead of CVCV). Indeed, the youngest child in the Auriemmo et al (2009) study was six years, three months with a hearing loss, which may have made speech articulation errors likely. Johnson (2000) found that accurate scoring of children’s responses to the ED-NST was facilitated by having them wear a lavalier microphone connected to a camcorder that delivered a high-quality audiovisual signal to a 19-inch video monitor. Although such elaborate arrangements may not be necessary for routine clinical administrations of the ED-NST, it would be desirable in research studies and at least merits more detail in the methods provided by Auriemmo et al (2009). Moreover, it is important that researchers report on examiner scoring reliability; most studies have done so on about 5% of randomly selected portions of their data sets (e.g., Edgerton and Danhauer, 1979; Danhauer et al, 1985). Auriemmo et al (2009) reported neither exactly how experimenters scored the ED-NST nor whether either intra- or interexaminer reliability was determined, which could raise questions about the validity of their data.

Another concern has to do with the presentation level of the NST. Auriemmo et al (2009) said they presented the ED-NST in a sound field at 30 and 50 dB HL input levels in order to compare vowel and consonant identification scores across aided conditions to determine the efficacy of LFT. The authors reported that the children rated the loudness of a meaningful passage (i.e., “Summer in Sweden” [Plant, 2006]) to ensure audibility and comfort. These procedures, however, do not ensure audibility for children’s maximum performance on the NST stimuli across aided conditions. In addition, the selection of these input levels precluded determining whether the favorable results observed were due to differences in audibility or signal processing (i.e., LFT or no LFT) across the aided conditions. It is well known that children require more spectral energy than adults to perform optimally on speech recognition tasks (e.g., Elliott et al, 1986). In addition, previous studies have strongly advocated obtaining performance-intensity (PI) functions by presenting the ED-NST at multiple levels or at least at 35–45 dB sensation level (SL) in reference to listeners’ speech recognition thresholds (i.e., SL re: SRT) to ensure adequate audibility for maximum performance so that valid comparisons can be made across listening conditions. Using 30 and 50 dB HL input levels across aided conditions without determining children’s SRTs and appropriate SLs for presentation of the stimuli could be problematic. That is, it is difficult to discern whether the significant differences found in phoneme recognition were due to additional cues provided by LFT or simply to audibility of the stimuli across conditions. In other words, the playing field across aided conditions must be leveled for audibility prior to comparing the discriminability of cues with and without LFT. We wonder how audibility was ensured across aided conditions so that discriminability of cues could be compared with and without LFT. This could be an issue in the first two paragraphs of page 302; here the authors discuss the lack of improvement noted with LFT at 30 dB HL (where the stimuli may have been inaudible to the children, which may speak more to the benefits of the training itself than to the processing strategy). Also, the authors stated in paragraph three on page 302 that the speech stimuli might not have shown benefits as well as in the real-life situations that were reflected on the environmental sounds survey. This would not be so surprising if the experimental speech stimuli were actually below the children’s audibility levels, whereas sounds in real-life situations might be louder. It is difficult to evaluate the results of the study accurately without sufficient details about the administration and scoring of the ED-NST. For example, on page 297, the authors stated that an ANOVA was used to test the significance of the three within-subject effects of level (30 or 50 dB HL) × aided conditions (six) × phoneme position (two, initial and medial) for consonant identification and vowel identification separately. This is confusing; certainly the CVVC stimuli of the ED-NST allow for assessing initial and medial positions of consonants but not for vowels that never occur in the initial position of these targets. They reported that position was significant for consonants but not for vowels. We believe that they meant to say, “phoneme position (two, initial and medial for consonants, or medial and final for vowels) identification separately,” which would have been more accurate.
Then readers would know what the authors meant when they later stated on page 297, “Consequently, the averaged consonant scores and the averaged vowel scores were reported in all subsequent figures.”

Fortunately, we are at a time in this profession when technology has advanced to the point that we can help persons with almost any type and configuration of hearing loss as indicated by those in Auriemmo et al’s (2009) study. These advancements have made it necessary to use more difficult measures like NST stimuli devoid of semantics to study nuances in benefits provided by one processing scheme over others. We commend Auriemmo and colleagues for recognizing the need to use NST stimuli in their investigation. Certainly, materials for this type of stimuli are readily available and do not have to be reinvented for clinical or research use. We also commend them for choosing the ED-NST (not the CUNY NST as incorrectly referenced) because it is well suited with its multiple equivalent randomizations of the two 25-item CVCV lists for measuring the types of perceptual judgments required in this study without having to worry about memory or familiarity effects. One of the advantages of using the ED-NST is that if listeners’ responses are phonetically transcribed by the examiner, especially when the phonemes are scored for consonants, then the resulting error data can easily be converted to confusion matrices to evaluate how they used or misperceived various perceptual features. This would be helpful when evaluating different processing strategies such as those used in the hearing aids of this study. Perhaps the authors did collect this valuable information and could use it in further rehabilitation of these children’s hearing losses. This would also be useful in assessing changes in the children’s speech perception and production as outcomes of the hearing aids and the aural rehabilitation provided over time. Indeed, the CUNY NST and the ED-NST are both perfectly good tools for conducting such assessments; either would have been quite suitable for this study. However, because they are different tests, we suggest that Auriemmo et al need to provide readers with more specifics about which test was actually used and further details about its administration and scoring, which are critical for evaluating the outcomes of their study.

As a final comment, one of the first things we teach our students about consuming or conducting research is the requirement of having a complete, clear, and accurate presentation of the methods section. We stress upon them that the methods section is critical to evaluating any study’s results and conclusions and that all readers should be able to replicate it exactly by following the methods provided by the authors. Unfortunately, readers can neither replicate nor fairly evaluate the results of Auriemmo et al’s study because of the missing and inaccurate information from their methods. As reviewers and authors, we all know how difficult it is to conduct and publish research and how details can sometimes be overlooked. However, this particular study had no less than eight authors and successfully progressed through the rigorous review process to gain publication in this journal. In that process, the manuscript had to have been read by at least three peer reviewers, editors, and publication assistants. In fact, the study’s overall worth was validated by the fact that Dr. Jerger based his editorial comments for that issue of the journal on it. Thus, it is likely that nearly a dozen sets of eyes (eight authors, one editor, and three reviewers) reviewed the manuscript and missed the points raised here. Ultimately, it is the authors’ responsibility to catch such errors and omissions.

Interestingly, we just noticed the online publication of another article on this topic in this journal by one of the authors (Kuk et al, 2009) that described the NST on page 6 and again propagated the same errors in an earlier study they cited by Kuk et al (2007) in addition to citing the Auriemmo et al (2009) study. Oddly, Kuk et al (2009) briefly described a new nonsense syllable test, which they referred to as the “ORCA-NST” and say they developed for their study, but they failed to provide any normative data or details for it either. They stated that details about the development of their ORCA-NST were reported in another article (Kuk et al, submitted). Unfortunately, saying that details were presented in a manuscript submitted to an unidentified journal (see references) that apparently has not yet been accepted for publication does nothing to help readers grasp the information needed to evaluate any of their results in that or in their subsequent studies, a recurring problem with these authors’ research. Kuk et al (2009) claimed that they needed a test with multiple randomizations that would allow them to make several presentations of the stimuli in different conditions, yet provided no data to support that their new test met that objective. Again, normative data are needed. We also noted that Kuk et al (2009) incorrectly cited the title of the Auriemmo et al (2009) article in their references, another indication that they need to proofread their work more carefully. Again, the Kuk et al (2009) article seems to have worked its way thorough the review process of this journal without these errors being caught. Apparently, there is a need to correct these errors so that authors, reviewers, and readers are not misled further by and that they do not propagate wrong information.

We hope that this letter is not taken as being critical or contrary to the authors’ work, but that it is a friendly suggestion that will cause them to re-visit their methods and provide more detail so that others can fairly evaluate and use their findings. As part of this process, we were recently informed by the journal staff that the authors of Kuk et al (2009) were prompted to revise and correct the errors noted in the online version of their article in response to our comments. We accessed the online version on September 23,
2009, and the errors were still present. In reading the final print version of that article in the September issue, we noted that the reference to the ED-NST was corrected; however, the authors continued to state that details of their new ORCA-NST are available in a submitted, but still unnamed, journal. The journal staff noted that online prepublications do not receive final copyediting or authors’ review of proofs, and that neither process had eliminated these errors in this case. This is a good point that readers should keep in mind when viewing online prepublications. The important thing is that the errors were caught and mostly corrected before print publication, showing that the system does work. This all just shows that research and publishing is indeed a team effort. Researchers conduct and write up their findings, peers and staff (not for content) review and publish the work, and consumers read it and provide feedback, which all promotes the scientific inquiry process.

Carole E. Johnson
Auburn University

Jeffrey L. Danhauer
University of California Santa Barbara

REFERENCES


