Hearing Research 287 (2012) 1–2
Contents lists available at SciVerse ScienceDirect
Hearing Research
journal homepage: www.elsevier.com/locate/heares

Letter to the Editor

Questionable reliability of the speech-evoked auditory brainstem response (sABR) in typically-developing children

Hornickel et al., (2011a) have recently examined the test–retest consistency of speech-evoked auditory brainstem responses (sABRs) in "typically-developing children" both in quiet and in the presence of an acoustic background noise (4-speaker speech babble, 10 dB signal-to-noise ratio; using a six formant speech syllable/da/ as the stimulus). Based on the correlations between the tests taken at two points-in-time, they conclude that "reliability estimates were generally good." They further conclude that the sABR may be an unique tool for the assessment of auditory-based communication skills in children. Clearly, the reliability of sABRs is an important issue based on the suggestions that: 1) this test be included as part of a battery for the diagnosis of central auditory processing disorders in children (Billet and Bellis, 2011), 2) it has a relationship to reading ability and music aptitude in children (Hornickel et al., 2012; Strait et al., 2011) and, 3) it may have predictive value in assessing reading ability and speech-in-noise perception in school-age children (Hornickel et al., 2011b).

Although there are differences in the recommendations of experts (see Charter, 2003), a common view is that for use in clinical assessment, test–retest correlations above 0.80 are considered good and those below 0.70 are considered unacceptable (Cicchetti, 1994). Examination of the test-retest data from Hornickel et al. (2011a; Table 1), reconstructed and shown below (Table 1), reveals that only 1 of their 37 results is above 0.80 and only 2 are above 0.70. Thus, 34/37 (92%) of these correlations fail to reach a level commonly considered to be acceptable by clinical standards.

Hornickel et al. (2011a) report that 21 of 37 correlations are significant. However for clinical assessment, it is the magnitude of the correlation which is the critical. It is not sufficient to show that a significant amount of the variance in test scores is common to individuals in both testing sessions. It is also necessary to show that a large percentage of this variance is consistent. However, the Spearman rank ordered correlation coefficient is not based on variance, so this measure is an ambiguous index of reliability.

The importance of the actual magnitude of test–retest correlations was illustrated in a simulation study conducted by McFarland and Cacace (2006). This analysis showed that on the basis of reliability alone, at 0.80 the probability of correct diagnosis is about as likely as misdiagnosis (false alarms plus misses). At 0.70, either type of error is as likely as a correct diagnosis. Thus, unless a considerable portion of the variance in test scores is repeatable, diagnostic error is very likely. It is also important to stress that in addition to reliability, diagnostic accuracy also depends on the issue of validity (Cacace and McFarland, 1998; McFarland and Cacace, in press).

Problems in diagnostic accuracy due to poor reliability are compounded when there are multiple opportunities for examinees to fail a test. The simulations by McFarland and Cacace (2006) also showed that, considering only 5 tests, false positives are greatly elevated unless some precaution is taken for considering these tests as a group, such as averaging test results. With 37 test results, the sABR provides many opportunities for false alarms. Clearly, the issue of multiple measures would need to be considered if this test is to be used clinically.

The reliability coefficients reported by Hornickel et al. (2011b) were based on Spearman's rank order correlation. Spearman's rho is equivalent to the Pearson correlation calculated over continuous variables that have been transformed as ranks (Smith, 1986). This transformation might be useful if the rank-ordered data have some desirable properties not shared with the raw data. However, Hornickel et al. (2011a) provide no rational for their selection of this statistic. In any case, the use of Spearman's statistic implies that the reliabilities reported apply to the clinical use of these measures as rank-ordered values. If this is the intent, then this requirement should be stated explicitly. Otherwise, the reported rank-ordered correlations are not appropriate as measures of reliability as this is not the metric to be used clinically.

Table 1
Reconstruction of test–retest data from Hornickel et al. (2011a, Table 1).

Reliability Spearman's rho		
Response Latencies (ms)	Quiet	Noise
Peak 9	0.123	−0.185
Trough 10	0.139	−0.154
Peak 42	**0.565**	**0.566**
Trough 43	**0.456**	**0.401**
Peak 52	**0.473**	**0.590**
Trough 53	**0.484**	0.305
Quiet-to-noise Phase Shift (π radians) Low Harmonics		0.355
Within session replicability	**0.664**	**0.667**
Amplitude signal-to-noise ratio	**0.752**	**0.601**
Spectral Encoding (μV)		
F0	**0.815**	**0.656**
H2	**0.662**	0.231
H3	0.319	−0.195
H4	0.339	−0.117
H5	**0.586**	0.328
H6	**0.510**	**0.429**
H7	**0.740**	0.336
H8	0.202	0.142
H9	**0.540**	**0.511**
H10	0.358	**0.598**

Significant reliability correlation coefficients are bolded ($p < 0.05$). These data were taken from Hornickel et al. (2011a; Table 1).

0378-5955/$ – see front matter
doi:10.1016/j.heares.2012.02.014

Hornickel et al. (2011a) identify several issues with their results that might lead to a low reliability. One of these is the long test–retest period (average of 1 year) between tests which is much longer than in typical studies of reliability. Another issue is the use of a population of children without communication problems, which could lead to restriction in the range of scores observed. Collection of norms in a population of children to which this test would be applied for clinical purposes would be informative, particularly if it were in a sample larger than the 26 used by Hornickel et al. (2011a) as estimates of reliability in a population of this limited size would not be considered accurate (Charter, 1999).

In summary, although Hornickel et al. (2011a) have reported interesting results concerning the stability of sABRs in children, these data should not be used as a rational for the clinical application of this procedure. The evidence provided by Hornickel et al. (2011a) suggests that this methodology does not yet have sufficient reliability to justify its use as a clinical/diagnostic procedure.

References

Billet, C.R., Bellis, T.J., 2011. The relationship between brainstem temporal processing and performance on tests of central auditory function in children with reading disorders. J. Speech Lang. Hear. Res. 54, 228–242.

Cacace, A.T., McFarland, D.J., 1998. Central auditory processing disorder in school age children: a critical review. J. Speech Lang. Hear. Res. 41, 355–373.

Charter, R.A., 1999. Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. J. Clin. Exp. Neuropsychol. 21, 559–566.

Charter, R.A., 2003. A breakdown of reliability coefficients by test type and reliability method, and the clinical implications of low reliability. J. Gen. Psychol. 130, 290–304.

Cicchetti, D., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6, 284–290.

Hornickel, J., Knowles, E., Kraus, N., 2011a. Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. Hear. Res. 284, 52–58.

Hornickel, J., Chandrasekaran, B., Zecker, S., Kraus, N., 2011b. Auditory brainstem measures predict reading and speech-in-noise perception in school-aged children. Behav. Brain Res. 216, 597–605.

Hornickel, J., Anderson, S., Skoe, E., Yi, H.G., Kraus, N., 2012. Subcortical representation of speech find structure relates to reading ability. NeuroReport 4, 6–9.

McFarland, D.J., Cacace, A.T., 2006. Current controversies in CAPD: from Procrustes' bed to Pandora's box. In: Parthasarathy, T.K. (Ed.), An Introduction to Auditory Processing Disorders in Children. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 247–263.

McFarland, D. J., Cacace, A. T. Establishing the construct validity of the auditory processing disorder (APD): application of psychometric theory to clinical practice. In: Goldfarb R. (Ed.). Translational Studies in Speech Pathology and Audiology: Essays Honoring Dr. Sadanand Singh. Plural Publishing, San Diego, Ca. (in press).

Smith, R.B., 1986. Some properties of Rho-b statistics. Qual. Quant. 20, 53–74.

Strait, D.L., Hornickel, J., Kraus, N., 2011. Subcortical processing of speech regularities underlies reading and music aptitude in children. Behav. Brain Funct. 17 (7), 44.

Dennis J. McFarland*
Laboratory of Neural Injury and Repair, The Wadsworth Center New York State Health Department, Empire Plaza, Albany, NY 12201-0509, USA

Anthony T. Cacace
Department of Communication Sciences & Disorders, Wayne State University, Detroit, MI, USA

* Corresponding author. Tel.: +1 518 486 2677; fax: +1 518 486 4910.
E-mail address: mcfarlan@wadsworth.org (D.J. McFarland)

10 February 2012
Available online 15 March 2012