



# Validation of a tablet-based assessment of auditory sensitivity for researchers

Kathryn Wiseman<sup>1</sup> · Jerry Slotkin<sup>2</sup> · Meredith Spratford<sup>1</sup> · Amberlee Haggerty<sup>1</sup> · Maggie Heusinkvelt<sup>1</sup> · Sandra Weintraub<sup>3</sup> · Richard Gershon<sup>4</sup> · Ryan McCreery<sup>1</sup>

Accepted: 13 July 2022 / Published online: 12 August 2022  
© The Psychonomic Society, Inc. 2022

## Abstract

Quantifying hearing acuity is increasingly important across a wide range of research areas in the behavioral and neurosciences. Scientists have relied on either self-reported hearing status or the availability of diagnostic hearing assessment in past studies. There remains a need for a valid and reliable assessment of auditory sensitivity that can provide estimates of the magnitude of hearing loss, if present, without requirements for professional audiologists, facilities, and equipment that are needed to conduct a diagnostic hearing assessment. The goal of this experiment was to validate the NIH Toolbox® Hearing Threshold Test (HTT), a tablet-based hearing assessment available via iPad application that uses consumer-grade headphones, on a clinical sample of children and adults with varying degrees of hearing acuity. Electroacoustic analysis of the hearing assessment application and headphones demonstrated acoustic outputs within established conformity standards for hearing assessment. Twenty-seven children and 63 adults participated in a standard diagnostic hearing assessment and the experimental tablet-based assessment. The results showed that thresholds from the tablet-based assessment were highly correlated with thresholds from the clinical hearing assessment ( $r = .83-.93$ ) for children and adults for all frequencies and across a range of levels of hearing acuity. The HTT also met clinical test–retest reliability standards (Cronbach's  $\alpha > .86$ ). The tablet-based hearing assessment provides acceptable estimates of hearing levels for children and adults when diagnostic audiometric assessment capabilities are not available.

**Keywords** Hearing assessment · Tablet · iPad · NIH Toolbox · Adults · Children · Hearing loss

## Introduction

Assessment of hearing sensitivity has become increasingly important in research of cognitive psychology and neuroscience. Individual differences in hearing sensitivity are known to affect language, cognitive, and academic outcomes in children (Kronenberger & Pisoni, 2020;

McCreery & Walker, 2021; Tomblin et al., 2015), even in cases with very subtle differences in sensitivity from the typical range (McCreery et al., 2020; Moore et al., 2020). For adults, hearing loss has been identified as a risk factor for dementia (Lin et al., 2011; Livingston et al., 2020; Thomson et al., 2017), depression (Lawrence et al., 2020) and as a contributing factor to social isolation (Dawes et al., 2015; Maharani et al., 2019). Even when hearing sensitivity is not a primary research question, there is a need for researchers to document the hearing status of participants when studying related domains of language, cognition, and behavior (Füllgrabe, 2020). Diagnostic assessment of hearing often is expensive (Abu-Ghanem et al., 2016; Yousuf Hussein et al., 2016), as it requires specialized equipment, space, and personnel. Though some researchers have access to these clinical assessments, they are not widely available in many research settings. Barriers to documenting hearing sensitivity in research studies has led many scientists to rely on self-reported

---

✉ Ryan McCreery  
ryan.mccreery@boystown.org

<sup>1</sup> The Audibility, Perception, and Cognition Laboratory, Center for Hearing Research, Boys Town Research Hospital, 555 North 30th Street, Omaha, NE 68131, USA

<sup>2</sup> University of Delaware, Newark, DE, USA

<sup>3</sup> Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern Feinberg School of Medicine, Chicago, IL, USA

<sup>4</sup> Department of Medical Social Sciences, Northwestern Feinberg School of Medicine, Chicago, IL, USA

hearing difficulty. Self-reported hearing assessments have limitations for predicting the magnitude of hearing loss, have yet to be validated with children, and have poorer sensitivity and specificity than pure-tone screening methods (Clark et al., 1991; Louw et al., 2018; Nondahl et al., 1998; Tsimpida et al., 2020). The purpose of this study, as part of the Advancing Reliable Measurement in Alzheimer's Disease and Cognitive Aging (ARMADA) study (Weintraub et al., 2021), was to evaluate the acoustic performance of and behavioral responses to a tablet-based measure of hearing sensitivity used with consumer-grade headphones. Compared to standard pure-tone audiometry completed in an audiology clinic, the tablet-based hearing assessment paired with consumer-grade headphones could address the need for more accessible hearing assessments in a broad range of research settings.

There are several types of hearing assessments that vary in specificity (screening, diagnostic) and methodology (self-report, pure-tone audiometry) of measurement. The audiological diagnosis of hearing loss relies, in part, on hearing assessment via pure-tone audiometry (Roeser, 2013). Pure-tone audiometry is a process that establishes absolute sensitivity for pure-tone signals across the range of frequencies important for understanding speech (ASHA, 2005; Brender et al., 2006). There are rigorous standards for room acoustics of the spaces where audiometric tests are conducted (Frank et al., 1993; Frank, 2000) and the specific equipment (e.g., diagnostic audiometers) and transducers needed to present the signals (ANSI S3.6 2018), all of which needs precise calibration (ANSI/ASA S3.1-1999 R2018; Champlin & Letowski, 2014). In addition to equipment, specialized healthcare professionals are needed to conduct diagnostic audiologic assessments. Audiologists, who diagnosis disorders of hearing and balance, have graduate- or doctoral-level training, are licensed to perform diagnostic audiometry, and typically work in medical, industrial, or educational settings (Martin & Clark, 2019). Because of these equipment and personnel requirements, pure-tone audiometry is often not feasible for research studies in neuro- and behavioral sciences outside of specialized hospitals or research centers (e.g., Amieva et al., 2015; Amieva et al., 2018; Vaccaro et al., 2019; Vassilaki et al., 2019; Zekveld et al., 2013).

The limited availability of pure-tone audiometry means that many scientists have relied on measures of self-reported hearing status when examining the effects of hearing loss on cognitive or behavioral outcomes or determining if a participant meets inclusion or exclusion criteria (Reilly et al., 2007). While self-reported hearing assessments are important for characterizing an individual's subjective, functional hearing, particularly related to the impact of their disability (e.g., limitations in activity and participation; ASHA, n.d.; World Health Organization, 2001), it is not an objective measure of hearing sensitivity and is a proxy measure for

an individual's impairment (e.g., physiological limitations; ASHA, n.d.; World Health Organization, 2001).

Measures of self-reported hearing loss have been shown to predict the presence or absence of hearing loss in healthy adults with reasonable accuracy (e.g., Clark et al., 1991; Nondahl et al., 1998; Salonen et al., 2011). However, the relationship between self-reported hearing status and the magnitude or degree of hearing loss for adults is less precise. In a study that pooled results from several large, public-health datasets for over 23,000 adults, Kiely et al. (2012) found that the accuracy of self-reported hearing difficulty for predicting at least a mild degree of hearing loss depended on the age of the participants. The authors found that self-report generally underestimated the presence of hearing loss for adults who were over 75 years of age. They reported self-reported hearing loss (as compared to pure-tone average of better hearing ear) had sensitivity that ranged from 62 to 85% and specificity that ranged from 41 to 85% across datasets. Likewise, accuracy of self-reported hearing loss varies across demographic characteristics like race, ethnicity, and listener sex (Kamil et al., 2015; Schnohr et al., 2019).

Several technological solutions have been developed recently to mitigate barriers to hearing assessment for researchers, using advances in computer-tablet technology. These solutions fall into two broad categories: hearing screening applications and diagnostic audiometry applications. Hearing screening applications assess pure-tone hearing thresholds and provide an overall pass or fail result for each ear based on established criteria for normal or near-normal hearing (Bright & Pallawela, 2016; Corona et al., 2020; Swanepoel et al., 2014). Like measures of self-reported hearing loss, hearing screening only provides an indication of presence or absence of hearing loss and does not quantify the magnitude of hearing loss, if present. Tablet- or smartphone-based diagnostic audiometric applications have been developed and validated to meet established diagnostic standards as an alternative to standard diagnostic pure-tone audiometry (Bright & Pallawela, 2016; Irace et al., 2021; Sandström et al., 2020; Szudek et al., 2012; Thompson et al., 2015; Yeung et al., 2015). For instance, Thompson et al. (2015) found that one diagnostic application provided hearing threshold estimates that were within 10 decibels (dB) of standard diagnostic pure-tone audiometry for adults. However, these commercial diagnostic applications currently cost over \$5000 United States dollars and, similar to diagnostic pure-tone audiometry equipment, still require annual calibration by an expert. These costs may limit their utility for scientists who want an affordable way to obtain a valid and reliable estimate of hearing sensitivity for their research.

Until the last decade, scientists often faced similar cost and access barriers when completing standardized assessments of vision, cognition, language, and emotional functioning for their research. The National Institutes of Health

Toolbox for the Assessment of Neurological and Behavioral Function, or NIH Toolbox®, was developed to address these barriers and provide researchers with access to a set of standardized tools that assess sensory and cognitive abilities (Gershon et al., 2010). The NIH Toolbox application (<https://apps.apple.com/us/app/nih-toolbox/id1002228307>) is available through the App Store of the Apple iPad (NIH & Northwestern University, 2021), and researchers can access these assessments with the purchase of a subscription through the app. The NIH Toolbox includes a validated assessment of speech recognition in noise using the “Words-In-Noise Test” (Zecker et al., 2013). Speech recognition in noise is an important predictor of auditory ability that often relates to communication difficulties faced by people with hearing loss in real-world environments (Eckert et al., 2017), but it does not provide an estimate of the presence or degree of hearing loss. The NIH Toolbox Hearing Threshold Test (HTT) was recently developed to provide an automated, cost-effective assessment of hearing sensitivity (air conduction thresholds) that can be conducted with a commercially available computer tablet and consumer-grade headphones (i.e., available commercially, more cost-effective than healthcare-grade equipment). This application helps fill the gap between self-report or screening of hearing loss and standard diagnostic audiologic assessment by an audiologist. While this tool has been previously described in the literature (Zecker et al., 2013), the HTT has not been validated for use in children and adults with a variety of hearing levels.

There are several methods of validation used to determine whether an assessment measures its intended construct. Criterion validity refers to a tool’s performance compared to the standard tool used to measure the intended construct. To prove good criterion validity, the HTT should be compared to the standard diagnostic audiometry testing commonly used to diagnose hearing loss in a clinical setting. Tablet-based hearing assessments currently on the market have demonstrated 80.1–95.8% accuracy within 10 dB of a standard clinical audiometer (Saliba et al., 2017; Sandström et al., 2020; Thompson et al., 2015), suggesting precedent for the validity of tools similar to the HTT. If the HTT produces estimates of hearing sensitivity that approximate those of the standard pure-tone audiologic assessment, this would indicate the HTT could be used to provide ear-specific information about hearing sensitivity that could be used to account for hearing in behavioral and neuroscience research. In addition to examining validity of the HTT, it is important to examine whether accuracy differs systematically by listener characteristics. Studies of self-reported hearing loss suggest that the accuracy of a single-question measure of self-reported hearing loss differs by listener age for adolescents (Schnohr et al. (2019) and adults (Choi et al., 2016). Accuracy of self-reported hearing loss also differs by degree of hearing loss, with accuracy increasing as degree

of hearing loss worsens (Feltner et al., 2021; Fredriksson et al., 2016). However, no studies of tablet-based hearing assessments have examined the impact of age and degree of hearing loss on the accuracy of these assessments relative to standard pure-tone audiometry.

Research on the accuracy of new assessments should address their sensitivity and specificity and their test–retest reliability. Sensitivity and specificity are often reported for self-report measures of hearing loss (sensitivity ranging from 62 to 85% and specificity ranging from 41 to 85% across studies) and other tablet-based measures (sensitivity ranging from 88 to 100% and specificity ranging from 58 to 96% across studies), suggesting varied ability of these tools to accurately identify individuals with hearing loss (Bright & Pallawela, 2016; Kiely et al., 2012; Saliba et al., 2017; Sandström et al., 2020). However, test–retest reliability is rarely reported in feasibility and validation studies of tablet-based hearing threshold assessments (Sandström et al., 2020). Proving good test–retest reliability of the HTT would provide additional evidence of the utility of this measure for research purposes. The purpose of this study was to assess the criterion validity and reliability of the HTT for a sample of children and adults who had varying levels of hearing sensitivity to determine if this assessment could be used to estimate hearing levels for individuals who participate in behavioral research studies. The following research questions (RQs) were addressed:

**RQ1a:** How do thresholds measured via the HTT compare to thresholds measured via standard pure-tone air-conduction audiometry?

Hypothesis: The HTT will produce hearing thresholds that are highly correlated to thresholds measured via standard pure-tone audiometry.

**RQ1b:** Does validity of the HTT vary depending on the age or hearing level of the listener?

Hypothesis: There may be differences in accuracy by listener age or hearing level, but these differences will not impact the overall validity of the HTT.

**RQ2:** What is the test–retest reliability of the HTT?

Hypothesis: The HTT will have good test–test reliability (Cronbach’s  $\alpha > .8$ ).

**RQ3:** What is the specificity and sensitivity of the HTT? Hypothesis: We predict good specificity and sensitivity (i.e., greater than 90%).

## Methods

### Participants

Ninety participants were recruited using a variety of methods (i.e., research database, email campaign, outpatient

audiology clinic). Participants included in this study met the following criteria: a) were 6 years of age or older and b) had hearing within normal limits (< 20 dB hearing level [HL] better-ear pure-tone average [BEPTA] at 500, 1000, 2000, and 4000 Hz; ANSI S3.6-2018 R-2018) or indicated mild-to-severe hearing loss (20–70 dB HL BEPTA) based on their clinical audiogram results (World Health Organization, 1991). Exclusion criteria included any otologic condition requiring medical management (e.g., otitis media, sudden or recent changes in hearing, vestibular complaints) or that required developmental modifications to the test protocol, including visual reinforcement or conditioned response per medical record and/or patient or caregiver report. Table 1 shows the age and hearing loss characteristics of the participants.

The final sample included 27 children (mean age = 9.2 years, standard deviation = 2.2, range = 6–15) and 63 adults (mean age = 42.1 years, standard deviation = 17.4, range = 18–76). The child group included 23 children with normal hearing and 4 children with hearing loss (mean BEPTA = 33.8 dB HL). The adult group included 44 adults with normal hearing and 19 adults with hearing loss (mean BEPTA = 29.1 dB HL). All participants categorized in the hearing loss group had bilateral hearing loss. Six participants had mild hearing loss (i.e., PTA of 20–40 dB HL) in one ear, and normal hearing (i.e., PTA < 20 dB HL) in the other, but were categorized as having normal hearing due to their better hearing ear. Two participants in the hearing loss group had asymmetry between ears (PTA difference of > 15 dB). All participants consented or assented to participate, and parents of children completed informed consent. Participants were compensated \$15/h for their participation. The Institutional Review Board of Boys Town National Research Hospital approved the study procedures prior to data collection.

**Equipment**

The study consisted of a standard pure-tone air conduction audiometric assessment and the experimental tablet-based hearing threshold assessment (i.e., HTT). Standard pure-tone

audiometric assessments were completed on a Grason Stadler Inc. GSI-61 2-channel diagnostic audiometer (Eden Prairie, MN) using either Telephonics TDH-49 circumaural headphones (Farmingdale, NY) or Etymotic Research ER-3A insert earphones (Elk Grove Village, IL). The experimental tablet-based hearing threshold assessment was completed with a 5th generation Apple iPad (Cupertino, CA) with Sennheiser HD 280 Pro circumaural headphones (Wedemark, Germany). Sennheiser HD 280 Pro headphones were selected for the behavioral study because of their widespread commercial availability, relatively low cost, and output and frequency responses that were adequate for the HTT. The HTT assessment was administered through the NIH Toolbox app available through the App Store for iPad (NIH & Northwestern University, 2021). Acoustic calibration for both hearing assessments (clinical audiometric assessment and HTT) was completed using a Larson Davis 831 sound-level meter (Depew, NY) with a Larson Davis 2559 1/2" microphone in a 6 cm<sup>3</sup> flat-plate coupler for circumaural headphones.

**Procedure**

Prior to the administration of the hearing assessment, an electroacoustic validation of the experimental tablet-based HTT was completed to ensure the accuracy of the intensity and frequency of pure-tone signals. The electroacoustic validation was initially completed with several different iPad models to ensure consistency across different models of the same generation, including iPad and iPad Air. The output of the iPad was set to 50% volume. Test frequencies of 500, 1000, 2000, and 4000 Hz were chosen because of their importance for hearing and communication based on published standards (ANSI S3.5-1997). The intensity range of the tablet-based assessment was verified from 10 dB sound pressure level (SPL) to 90 dB SPL. Based on American National Standards Institute (ANSI) Calibration Standards for Audiometers, the output of the device was verified to be within 3 dB across the intensity range at each test frequency to ensure accuracy and linearity of the audio output and headphones (i.e., Sennheiser HD 280 Pro headphones).

**Table 1** Participant characteristics

	Adult		Child		Overall	
	Normal Hearing	Hearing Loss	Normal Hearing	Hearing Loss	Normal Hearing	Hearing Loss
<i>n</i>	44	19	23	4	67	23
Age (years)	<i>M</i> = 42.1 <i>SD</i> = 17.4 Range 18–76	<i>M</i> = 56.7 <i>SD</i> = 11.4 Range 34–70	<i>M</i> = 9.2 <i>SD</i> = 2.2 Range 6–15	<i>M</i> = 12 <i>SD</i> = 1.1 Range 11–13	<i>M</i> = 30.8 <i>SD</i> = 21.1 Range 6–76	<i>M</i> = 48.9 <i>SD</i> = 20.0 Range 11–70
BEPTA (dB HL)	<i>M</i> = 8.8 <i>SD</i> = 5.8 Range 0–18.8	<i>M</i> = 29.1 <i>SD</i> = 8.8 Range 20–50	<i>M</i> = 3.5 <i>SD</i> = 2.2 Range 0–7.5	<i>M</i> = 33.8 <i>SD</i> = 9.4 Range 21.3–43.8	<i>M</i> = 7.0 <i>SD</i> = 5.5 Range 0–18.8	<i>M</i> = 29.9 <i>SD</i> = 9.0 Range 20–50

BEPTA = Better-ear pure-tone average; dB HL = Decibels hearing level

The behavioral validation of the experimental tablet-based hearing threshold assessment consisted of a diagnostic pure-tone air conduction audiogram and the HTT for each participant. For both the clinical audiogram and the HTT, each participant was seated in a large, sound-treated audiometric test booth. The diagnostic clinical audiogram was completed first to ensure that each participant had thresholds within the range that could be assessed by the experimental assessment. For both hearing tests, participants were instructed to respond when they heard the tone, even if it was perceived to be soft or barely audible. Response feedback was not provided for either the diagnostic or experimental hearing test.

For the standard diagnostic audiogram, circumaural headphones or insert earphones were placed on the participant's ears by the research assistant, and the participant was given a response button to press when they heard the sound. Air conduction testing was performed using pure-tone signals via a modified method of limits (Hughson & Westlake, 1944), where the audiologist decreased the level of the signal by 10 dB HL when the participant responded to the signal and increased the level of the signal by 5 dB HL when the participant failed to respond to the signal. Threshold was defined as the lowest level in dB HL where the participant responded to two-thirds consecutive ascending trials (Carhart & Jerger, 1959; Hughson & Westlake, 1944; Jerger, 2018). As in standard clinical practice, thresholds could be tested within the limits of the audiometer down to  $-10$  dB HL and up to 110 dB HL. Thresholds to pure tones were measured via air conduction at octave frequencies from 500 to 4000 Hz in each ear. To match the HTT protocol, bone conduction testing was not performed. If the participant had completed a standard diagnostic audiogram in the clinic within the past 6 months, those pure-tone threshold results were used in this study, as previous data show limited changes in sensorineural hearing loss for children (McCreery et al., 2015) and adults (Cohn, 1999) within this timeframe.

For the experimental, tablet-based HTT, consumer-grade Sennheiser HD 280 Pro headphones were placed on the participant's ears by the research assistant, and the tablet was placed in front of the participant on a stand, with the screen facing the participant. The participant was instructed to press a target area/button on the screen of the tablet when they heard a sound. A single practice tone (randomized; 500, 1000, 2000, or 4000 Hz) was administered at 70 dB SPL in either right or left ear (randomized) to familiarize the participants with the test stimuli. The experimental, tablet-based HTT used an automated, adaptive tracking procedure that followed a similar method of limits as the diagnostic audiometric assessment conducted by the audiologist. The experimental assessment started by presenting a signal at a specific frequency in one ear, randomized by frequency and ear for each participant. The initial presentation level began

at 60 dB SPL for each frequency and ear. If the participant responded to the signal at that level, the level of the signal decreased on the next trial by 20 dB SPL. If the participant failed to respond to the initial trial, the level of the signal increased by 5 dB SPL. After the initial presentations, the level of the signal was increased or decreased in the same manner as in a diagnostic audiogram based on each participant's response. The trials were administered in blocks by frequency and ear so that thresholds could be established within a block of trials (Harrell, 2002). The order of frequencies tested and test ear for each block of trials was randomly selected for each participant from 500, 1000, 2000, and 4000 Hz. If the tracking procedure reached the limits of the intensity range for a specific frequency, three consecutive responses (at the lower limit 0 dB SPL) or three consecutive non-responses (at the upper limit of 90 dB SPL) led to the limit being established as the threshold level. The end of a block occurred after five reversals or, in the cases where a limit was reached, after the participant responded to the final presentation of the block. The threshold was calculated as the arithmetic mean of the last five reversals and converted from dB SPL into dB HL by the application. A subset of 76 participants (53 adults, 23 children) completed the experimental test a second time after a short break to assess the test–retest reliability of the experimental test.

## Statistical method

All statistical analyses and data visualizations were completed using the R Statistical Computing Language (version 4.0.3, R Core Team, 2021; packages: ggplot2, lem4, BlandAltmanLeh). To conduct electroacoustic validation, acoustic output of the HTT was compared to the indicated level in the HTT application for each frequency. Descriptive statistics were calculated for each frequency, and Pearson correlations were calculated between measured output and the indicated level. First, descriptive statistics were calculated for each test method and test frequency. To address RQ1a, we quantified the difference in thresholds by test methods using root-mean-square (RMS) error. RMS error was calculated by taking the geometric mean of the differences between the HTT and the audiometer threshold at each frequency for each individual. In addition, thresholds from the standard diagnostic audiogram and tablet-based hearing assessment were compared using a Bland–Altman analysis (Bland & Altman, 1999) that quantified the magnitude of errors at each frequency, with the diagnostic audiogram as the clinical standard and the experimental test as the comparison.

To address RQ1b, a linear mixed effects model with a random intercept for each participant was used to assess the effects of the participant's age and threshold level (dB HL) on the magnitude of the difference between the diagnostic audiogram and the HTT with test frequency and ear as a

repeated measure within each participant. We include data from both right and left ear for each participant in this model because it provides a larger dataset; therefore, we accounted for dependence between ears by including a random intercept for each participant, signaling to the model that multiple measurements were taken from the same person.

To address RQ2, the test–retest reliability of the experimental test was assessed using Cronbach’s  $\alpha$ . The clinical standards for test–retest reliability for behavioral audiometry of  $\pm 7$  dB for adults and children over 13 years of age (Mahomed et al., 2013) and  $\pm 10$  dB for children under 13 years of age (Beahan et al., 2012) were used as boundaries for clinically significant differences (RMS error) at each frequency. To address RQ3, test sensitivity and specificity of the HTT relative to the diagnostic audiological assessment were calculated for adults and children using a 2x2 confusion matrix of predicted versus true outcomes. This was calculated based on the classification of either normal hearing (BEPTA < 20 dB HL) or hearing loss (BEPTA  $\geq$  20 dB HL), based on each participant’s BEPTA measured via the HTT compared to the BEPTA from the diagnostic audiometric assessment. We calculated sensitivity by dividing the number of true positives (BEPTA  $\geq$  20 dB HL per diagnostic audiometric assessment and HTT) by the number of participants with hearing loss per diagnostic audiometric assessment. We calculated specificity by dividing the number of true negatives (BEPTA < 20 dB HL per diagnostic audiometric assessment and HTT) by the number of participants with normal hearing per diagnostic audiometric assessment.

## Results

### Electroacoustic analysis

The acoustic output of the HTT for each test frequency measured by the sound-level meter in dB SPL was compared to the indicated level in the HTT application at each test frequency for a range of inputs from 35 dB SPL to 90 dB SPL. The accuracy and linearity of the calibration was assumed by analyzing the absolute difference in dB and Pearson correlation between the measured levels and the levels indicated in the HTT application. The mean difference across frequencies was 0.49 dB (means: 500 Hz = 0.57 dB, 1000 Hz = 0.32 dB, 2000 Hz = 0.5 dB, 4000 Hz = 0.66 dB). The maximum difference at any frequency was 1.5 dB. The correlation between the HTT-indicated level and measured level was  $r = 0.99$  across frequencies, indicating near perfect correspondence and consistent with the small differences between the indicated and measured levels in the application.

### Behavioral validation

The average thresholds for each ear of children and adults grouped by hearing status for the standard diagnostic audiometric assessment and the HTT are shown in Table 2. The average test time for the HTT was 4.9 min ( $SD = 1.3$ ) for children and 5.0 min ( $SD = 1.4$ ) for adults. Individuals with hearing loss ( $M = 6.0$  min,  $SD = 2.8$ ) took 1.4 min longer to complete the test than participants with normal hearing ( $M = 4.6$  min,  $SD = 1.2$ ) on average.

Figure 1 shows the HTT thresholds plotted against those from the standard diagnostic audiometric assessment, with a different panel for each test frequency. For individuals with two HTT threshold measurements (for reliability), the HTT threshold closest to the diagnostic audiometric threshold (i.e., smallest difference) was taken for comparison to their audiometric threshold. The closest threshold was taken because the close correspondence between HTT tests was high. The Pearson correlation between the best HTT threshold and diagnostic audiometric thresholds ranged from .80–.93 across frequency, suggesting strong correspondence between the two measures.

Table 3 shows the RMS error between standard diagnostic audiometry and HTT by age group and hearing status. The proportion of each group with a clinically significant difference ( $\pm 7$  dB for adults;  $\pm 10$  dB for children) between the two measures is also shown. The mean differences and RMS errors for adults were smaller than for children, but all mean values were within current clinical standards for accuracy for diagnostic audiometry for each age group. These results suggest that the classification of degree of hearing loss for the HTT would be similar to that provided by diagnostic audiometric assessment.

Bland–Altman analysis was used to assess the agreement between the HTT and the diagnostic audiometric assessment. Figure 2 shows the difference between the HTT and the diagnostic audiometric assessment as a function of the average of the two measures for each participant. A difference of zero indicates agreement between the HTT and diagnostic audiometric assessment. The errors were within acceptable limits and did vary as a function of the average of the two measures across frequency.

To address RQ1b, we examined whether the validity of HTT varies by listener characteristics using linear mixed effects model (Table 4) to examine the effects of listener age and degree of hearing loss on the RMS error for each frequency for the HTT. The main effect of age indicated that younger participants had larger errors, as expected. The effect of pure-tone average was not significant and had a small coefficient, suggesting that the RMS error did not systematically vary as a function of degree of hearing loss. The main effect of age and nonsignificant effect of pure-tone average should be interpreted considering the significant age by pure-tone

**Table 2** Hearing thresholds (dB HL) for audiometry and HTT by frequency and group

	Adult		Child	
	Normal Hearing Mean (SD)	Hearing Loss Mean (SD)	Normal Hearing Mean (SD)	Hearing Loss Mean (SD)
500 Hz Audiometry	9.66 (6.77)	22.24 (14.03)	7.28 (5.34)	25.63 (10.84)
500 Hz HTT	8.19 (4.26)	16.05 (11.96)	8.46 (4.03)	26.75 (16.10)
1000 Hz Audiometry	8.24 (7.27)	25.79 (14.50)	3.15 (3.05)	33.75 (16.85)
1000 Hz HTT	12.00 (3.18)	24.32 (15.08)	11.61 (3.39)	34.38 (17.12)
2000 Hz Audiometry	9.77 (8.09)	33.16 (13.82)	4.13 (4.63)	41.25 (16.42)
2000 Hz HTT	14.68 (6.39)	32.92 (14.39)	13.91 (4.31)	41.63 (21.40)
4000 Hz Audiometry	14.60 (13.52)	48.68 (14.03)	2.50 (3.46)	46.88 (11.00)
4000 Hz HTT	16.99 (12.35)	47.76 (14.22)	11.65 (8.70)	51.38 (11.19)

dB HL = Decibel hearing level; HTT = Hearing Threshold Test; Hz = Hertz; Participants were grouped by hearing status according to better-ear pure-tone average; therefore, specific thresholds may better or worse than the definition of hearing loss (threshold  $\geq 20$  dB HL). Thresholds for both ears are included for each participant

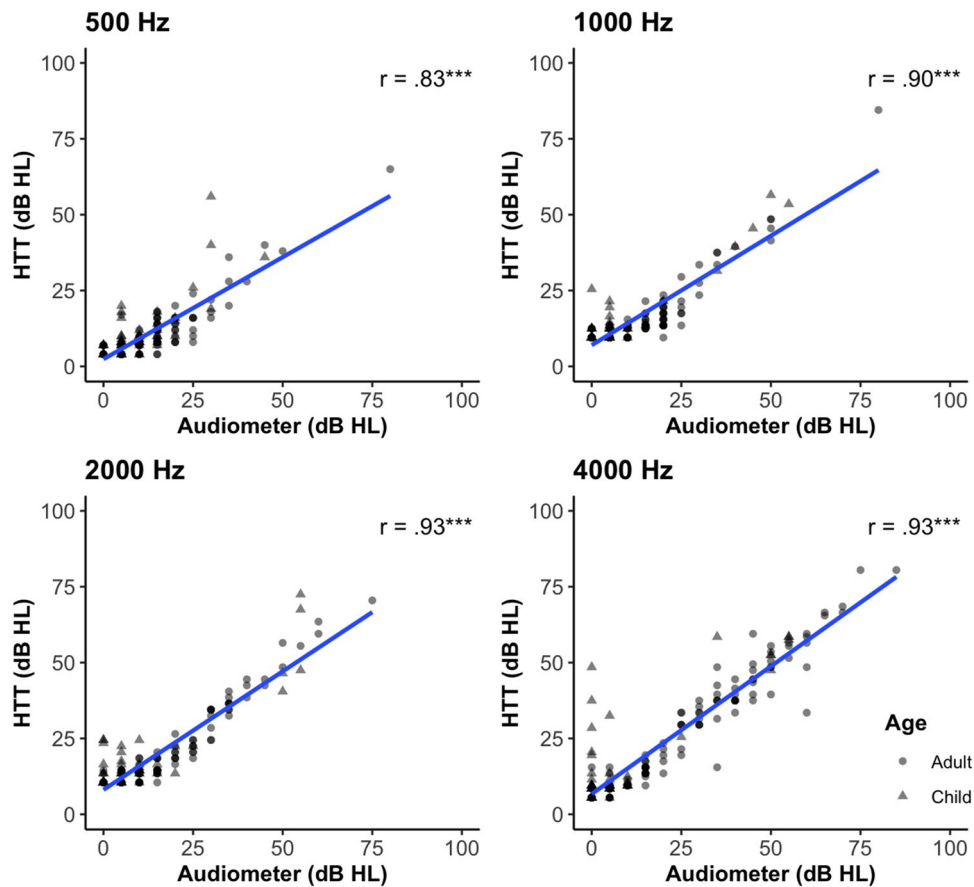
average interaction, which suggested that the RMS error increase with pure-tone average was slightly greater in the children than in the adults. There were no significant differences in RMS error between ears of the same participants. The frequency-specific effects suggest that while the RMS error was acceptably low across all test frequencies, the RMS error was significantly lower at 500 Hz than at 1000 Hz and 2000 Hz. The statistically significant differences in RMS error across frequency were not clinically significant.

To address RQ2, we examined the test–retest reliability of the HTT using a subset of participants ( $n = 76$ ) who completed two HTT assessments at the same visit. Figure 3 shows the 1st HTT plotted against the 2nd HTT for each participant. A reliability run for one participant was removed due to observed inattention during the task. The Cronbach's  $\alpha$  for each frequency was high, ranging from .87 to .97. These values indicate that the results were consistent across both HTT assessments for each participant. Table 5 shows the RMS error between the two HTT assessments for each participant, the mean difference, and the proportion of each group with a clinically significant difference across HTT. Children had larger errors and a higher proportion of clinically significant differences than adults, but the RMS error and mean differences were within acceptable limits for audiometric diagnostic standards for test–retest reliability. Examining RQ3, the overall sensitivity for the HTT was 96% (22/23 participants with hearing loss correctly identified as hearing loss) and the specificity was 96% (64/67 participants with normal hearing correctly identified as having normal hearing).

## Discussion

This study evaluated the validity and reliability of a new assessment of hearing sensitivity, known as the Hearing Threshold Test (HTT), in the NIH Toolbox. Electroacoustic validation of the measured and intended output of the HTT were assessed, followed by an examination of criterion validity of the HTT compared to the standard diagnostic pure-tone audiological assessment. The electroacoustic validation indicated that the application, coupled to consumer-grade headphones, met the standards for accuracy and linearity that are established for diagnostic audiological assessment (ANSI S3.6-2018). The maximum tolerance of the HTT at any frequency was 1.5 dB, meeting the ANSI standard for audiometric calibration, which requires the recorded level to be within 3 dB of the indicated level. These results suggest that the application is sufficiently calibrated for use with Sennheiser HD 280 Pro circumaural headphones and an iPad. It is recommended that users perform regular biological checks with this test and perform recalibration if irregularities are noticed.

The behavioral validation study indicated that the HTT produces results comparable to standard diagnostic audiometry, as differences between the HTT and diagnostic audiometric assessment were generally within test–retest reliability standards for the adults ( $\pm 7$  dB; Mahomed et al., 2013) and children ( $\pm 10$  dB; Beahan et al., 2012). These results provide evidence of good criterion validity of the HTT. The high sensitivity (96%) and specificity (96%) of the HTT demonstrate its ability to accurately identify the presence of



**Fig. 1** Relationship between thresholds measured via audiometer and the Hearing Threshold Test app. \*\*\* $p < .001$ ; HTT = Hearing Threshold Test; thresholds from the HTT were repeated (i.e., two runs). HTT thresholds plotted here represent the closest thresholds (i.e., smallest difference) to the audiometer thresholds. The blue line

represents the linear relationship between the variables. Individual thresholds are represented by gray points (circles for adults and triangles for children). Darker points indicate that multiple participants had the same thresholds

hearing loss in an individual. Overall, these results suggest that the HTT of the NIH Toolbox can provide researchers with efficient, accurate estimates of hearing sensitivity for research purposes at a considerably lower cost than a diagnostic audiometric assessment. At the time of this study, the cost of the application, iPad, and headphones was under

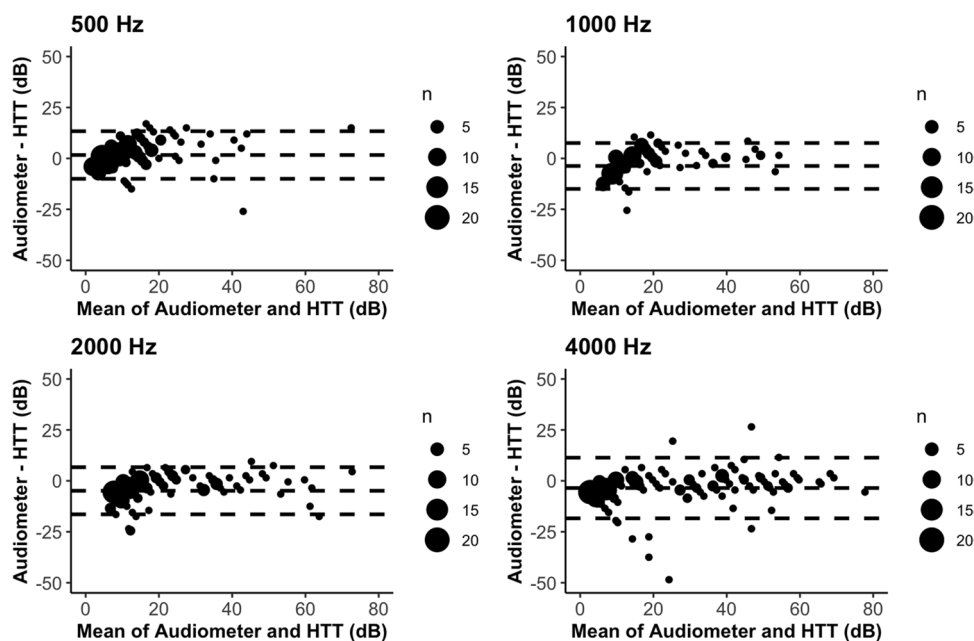
\$1000, making it more financially accessible than other currently available audiometers. The tablet and headphones can be used for additional research tasks if needed, making the equipment more efficient for use in a research environment. The HTT addresses the current limitations of screening and diagnostic tablet-based applications (e.g., need for

**Table 3** Root-mean-square (RMS) error in dB between audiometer and HTT thresholds by group and frequency

	Adult ( $n = 63$ )			Child ( $n = 27$ )			Overall		
	Clinically significant difference	RMS error (dB) <i>Mean (SD)</i>	Difference (dB) <i>Mean (SD)</i>	Clinically significant difference	RMS error (dB) <i>Mean (SD)</i>	Difference (dB) <i>Mean (SD)</i>	Clinically significant difference	RMS error (dB) <i>Mean (SD)</i>	Difference (dB) <i>Mean (SD)</i>
500 Hz	18.3%	4.59 (3.92)	2.89 (5.31)	9.3%	4.72 (4.50)	-1.17 (6.45)	15.6%	4.63 (4.09)	1.67 (5.96)
1000 Hz	27.8%	4.86 (3.21)	-2.18 (5.41)	13.0%	7.57(4.42)	-7.30 (4.87)	23.3%	5.67 (3.81)	-3.72 (5.75)
2000 Hz	22.2%	4.48 (3.43)	-3.36 (4.55)	53.7%	9.54 (5.45)	-8.39 (7.12)	31.7%	6.00 (4.73)	-4.87 (5.90)
4000 Hz	13.5%	4.02 (3.90)	-1.39 (5.44)	16.7%	8.61 (9.32)	-8.46 (9.45)	14.4%	5.40 (6.38)	-3.51 (7.59)

HTT = Hearing Threshold Test; Hz = Hertz; Difference = Audiometer threshold – HTT threshold; Clinically significant difference was defined as  $\pm 7$  dB for adults and children over 13 years of age (Mahomed et al., 2013) and  $\pm 10$  dB for children under 13 years of age (Beahan et al., 2012)





**Fig. 2** Bland–Altman analysis of difference between audiometer and HTT thresholds by frequency. dB HL = Decibel hearing level; HTT = Hearing Threshold Test; *x*-axis represents mean of the audiometer and HTT measurements, and the *y*-axis is the difference between

them. The *three dotted lines* represent the mean of the differences (center line) and 2 standard deviations above and below that. *Points* are scaled by number of occurrences of the value

trained personnel, lack of validation on children) and has sufficient accuracy to assess the magnitude of hearing loss (e.g., degree of hearing loss, PTA), if present.

**Validity and test–retest reliability of HTT**

This study tested whether the HTT produced estimates of hearing thresholds comparable to those obtained using standard diagnostic audiometry on a sample of children and adults who had either normal hearing or varying degrees of hearing loss. We found that the HTT reported thresholds that

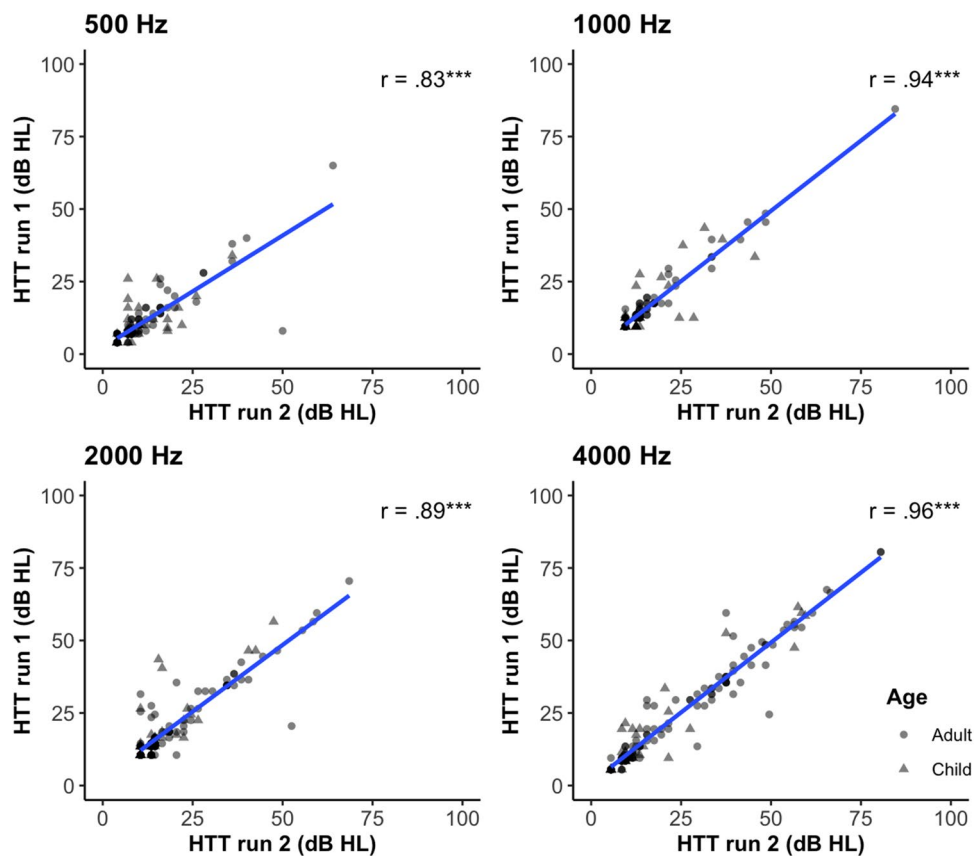
were, on average, within 5 dB or less for adults and within 9 dB or less for children compared to diagnostic audiometry. The current study showed comparable accuracy of threshold measurements relative to other validated tablet-based diagnostic applications currently on the market. These diagnostic applications have demonstrated 80.1–95.8% accuracy within 10 dB of a standard clinical audiometer (Saliba et al., 2017; Sandström et al., 2020; Thompson et al., 2015). While the current study used a stricter definition of accuracy ( $\pm 7$  dB for adults in the current study) to reflect current clinical standards for diagnostic audiological assessment, we found similar accuracy (87.5%) to other applications when using a 10 dB definition.

Additional findings from the current study differentiate the HTT from other applications. For instance, the accuracy of the HTT is notable given the inclusion of children and adults with and without hearing loss. Most other validated tablet-based diagnostic applications on the market have not been validated using children, though there are a few exceptions (Thompson et al., 2015; Yeung et al., 2015). Validation studies that include children either focus on applications specifically designed for younger children using conditioned play audiometry methodologies (Yeung et al., 2015) or include few children and did not perform separate analyses by age group (i.e., children and adults; Thompson et al., 2015). The HTT was shown to be valid measure of individuals ages 6 years or older, making it suitable for testing hearing sensitivity in both adults and children. Furthermore,

**Table 4** Linear mixed effects model examining effects of age and hearing status on RMS error by frequency

	Estimate	St. Error	t (df)	p value
Intercept	7.62	0.76	10.05 (132.5)	< .001
Age (years)	−0.09	0.02	−4.42 (103.1)	< .001
PTA (dB HL)	0.08	0.04	−1.83 (109.4)	.07
Age*PTA	0.002	0.001	2.24 (106.5)	.03
Frequency (reference = 500 Hz)				
1000 Hz	1.04	0.43	2.41 (625.4)	.02
2000 Hz	1.37	0.43	3.17 (625.4)	.002
4000 Hz	0.77	0.77	1.78 (625.4)	.08
Ear (Left vs. Right)	0.09	0.31	0.28 (633.4)	.78

dB HL = Decibel hearing level; Hz = Hertz; PTA = Pure-tone average (of each ear); RMS = Root-mean-square



**Fig. 3** Relationship between repeated HTT thresholds (run 1 vs. run 2). \*\*\* $p < .001$ ; dB HL = Decibel hearing level; HTT = Hearing Threshold Test; The blue line represents the linear relationship

between the variables. Individual thresholds are represented by gray points (circles for adults and triangles for children). Darker points indicate that multiple participants had the same thresholds

the accuracy of the HTT is particularly noteworthy, given that the application uses a commercially available tablet and consumer-grade headphones, rather than the tightly controlled equipment and transducers of clinical assessments (ANSI/ASA S3.1-1999 R2018; ANSI S3.6 2018; Champlin & Letowski, 2014).

The sensitivity and specificity of the HTT for classifying the presence or absence of hearing loss was high. Other studies of commercially available tablet-based systems

have shown mixed results with respect to sensitivity and specificity. For instance, validation studies of one application demonstrate good sensitivity (> 98% across studies), but variable specificity (ranging from 60.0 to 82.1% across studies). While results suggest the validity of this application to screen for hearing loss, a few studies concluded that it should not be used to obtain threshold-specific estimates of hearing sensitivity (Abu-Ghanem et al., 2016; Bright & Pallawela, 2016; Peer & Fagan, 2015; Szudek et al., 2012).

**Table 5** Test–retest differences for the HTT by frequency and age group

	Adult (n = 53)			Child (n = 23)			Overall		
	Clinically significant difference	RMS error (dB) Mean (SD)	Difference (dB) Mean (SD)	Clinically significant difference	RMS error (dB) Mean (SD)	Difference (dB) Mean (SD)	Clinically significant difference	RMS error (dB) Mean (SD)	Difference (dB) Mean (SD)
<b>500 Hz</b>	3.8%	1.75 (4.36)	−0.10 (4.70)	8.7%	3.09 (4.30)	0.13 (5.31)	5.3%	2.15 (4.37)	−0.03 (4.88)
<b>1000 Hz</b>	0.9%	0.93 (1.62)	0.48 (1.81)	15.2%	3.02 (4.48)	0.37 (5.41)	5.3%	1.57 (2.96)	0.45 (3.32)
<b>2000 Hz</b>	7.5%	2.32 (4.54)	0.49 (5.08)	6.5%	3.00 (5.79)	2.13 (6.17)	7.2%	2.53 (4.94)	0.99 (5.46)
<b>4000 Hz</b>	7.5%	2.42 (4.16)	0.03 (4.82)	10.9%	2.93 (4.12)	0.89 (4.99)	8.6%	2.58 (4.14)	0.29 (4.87)

dB = Decibel; HTT = Hearing Threshold Test; Hz = Hertz; Difference = Audiometer threshold – HTT threshold

This study also examined the test–retest reliability of the HTT by comparing two administrations of the tablet-based assessment within the same individual. The high reliability (Cronbach's  $\alpha > .86$ ), high correlations ( $r > .80$ ), and small differences between test administrations (mean  $< 3$  dB) suggest that this measure provides consistent, stable estimates of hearing from one test administration to the next. These results suggest that any discrepancies in magnitude of hearing loss over time most likely reflect changes in hearing rather than quality of the measurement. Most tablet-based, threshold-specific apps that are currently on the market have not reported their test–retest reliability. In one exception, Sandström et al. (2020) reported good test–retest reliability of a tablet-based application, but they only retested a single frequency (1000 Hz) and the retest occurred within the same test administration. The good test–retest reliability of the HTT provides additional evidence of the utility of this measure for research purposes.

Overall, evidence of strong validity, high sensitivity and specificity, and good reliability of the HTT suggests that this test can be used by researchers as an alternative to diagnostic audiological assessment to determine the presence and/or degree of hearing loss without specialized equipment or personnel. The accurate categorization of participants by hearing status (i.e., normal hearing versus hearing loss) and quantification of degree of hearing loss (e.g., BEPTA) can allow researchers to account for the participants' hearing levels when researching systems complementary to audition, such as language, cognition, memory, and psychosocial functioning. For instance, previous studies suggest that participants with hearing loss may have greater difficulty hearing and comprehending stimuli in tasks involving spoken language (e.g., word learning; Pittman et al., 2005; Stiles et al., 2013) or understanding verbal task instructions (Füllgrabe, 2020). Tools like the HTT allow researchers to account for hearing difficulties as they relate to their area of interest (e.g., language development, memory). Characterizing a participant's hearing status also allows researchers a more controlled approach to confirming eligibility for a study, if the presence of hearing loss is a criterion for inclusion or exclusion.

### Factors related to differences by test method

This study examined whether the listener's age or degree of hearing loss had an impact on the accuracy of the HTT, which has yet to be examined in studies of tablet-based assessments. Any effects of these factors on the accuracy of the HTT could have implications for the potential range of ages or degrees of hearing loss where the test could be used. Larger differences between the HTT and diagnostic audiometric assessment for listeners with normal hearing were observed for children and for people with better hearing

thresholds than adults or individuals with poorer hearing thresholds. Children are known to have greater variability in their estimates of hearing thresholds than adults (Beahan et al., 2012) due to a range of factors, including greater likelihood of lapses in attention and the inability to suppress self-generated noise during the assessment (Buss et al., 2016). These inattentive or noisy trials are often caught and discarded when testing is conducted by a clinician, but this is not the case for the automated HTT, which may help explain these larger differences in children. The effect of age on HTT accuracy showed an interaction with degree of hearing loss, suggesting that discrepancies between the HTT and the standard diagnostic audiogram are more likely to be larger in younger children with normal hearing, relative to other individuals. The larger differences between the HTT and diagnostic audiometric assessment in young individuals with better hearing thresholds were likely related to the fact that diagnostic assessments test at much lower sound levels for individuals with normal hearing levels than does the HTT, which are more easily masked by self-generated noise. As a result, some individuals with normal hearing had higher thresholds on the HTT, but these differences did not have an impact on the accuracy of the application.

### Limitations

Although the HTT provided highly accurate estimates of hearing sensitivity for children and adults in this study, there are several limitations to the application and the study results that should be considered. The HTT is not designed as a replacement for diagnostic audiometry. The HTT provided estimates of hearing sensitivity that were within established clinical standards for test–retest reliability and accuracy, but the HTT is not intended to diagnose specific hearing problems or replace diagnostic audiological assessment by an audiologist. It is meant to be used as a research tool. Medical diagnosis of hearing problems includes full diagnostic audiological assessment that includes pure-tone testing via air-conduction through headphones and via bone conduction to confirm potential sites of lesion in the auditory pathway. The HTT does not include bone conduction testing and as a result is not able to differentiate between different types of hearing loss, including conductive, sensorineural, or mixed. Additionally, the HTT cannot test hearing levels  $\geq 90$  dB SPL due to equipment constraints, limiting its utility in assessing profound degrees of hearing loss. The HTT does not require professional calibration, but the current data cannot demonstrate the stability of this system over time or substantiate the need for periodic calibration checks beyond biological calibrations and listener checks by end users.

This application has limitations in measuring thresholds in individuals with unilateral or asymmetric hearing loss, as this application does not use masking to prevent crossover

of loud stimuli to the contralateral ear. The HTT also does not test at higher frequencies (e.g., 6000, 8000 Hz) and is therefore limited in identifying listeners with high-frequency hearing loss, which is more commonly found in older individuals and individuals with noise induced hearing loss. The sensitivity and specificity for the HTT were high and support the application of these results for establishing the presence or absence and degree of hearing loss in children and adults, but the number of participants in this study with hearing loss was relatively small and further study of the HTT could provide a larger normative sample with hearing loss of varying characteristics.

Further, the results suggest that the HTT is an effective assessment for children, but this study did not assess children younger than school age. This assessment may not be applicable to young children, such as those who require alternative testing procedure (e.g., conditioned play audiometry) to complete an audiological assessment. Some children may require specific adaptations or closer monitoring by the test administrator to maintain their interest and attention in the HTT. Another limitation of the current study was that the test environment for the study was a quiet audiometric test booth in an audiology clinic. While previous work suggests minimal differences in hearing thresholds measured in a sound booth compared to a quiet room (Foulad et al., 2013), results for assessment in a less acoustically controlled context might limit the ability of the test to differentiate normal hearing from mild degrees of hearing loss in realistic settings. Any hearing assessment is only as effective as the acoustic environment where the test can be completed, and the results of the HTT obtained outside of the controlled environment tested here should be interpreted in the context of the ambient noise levels in the space. Headphones that cover the ears (e.g., circumaural) can provide some sound isolation but are not a replacement for attempts to minimize sources of ambient noise in any environment where hearing assessment will be conducted. Test administrators should monitor the ambient noise in the testing environment (e.g., sound level meter) to ensure quiet conditions and adjust the environment if noise will interfere with testing. Ideas for future development of this application could include adding features that would be useful for testing in non-standard environments with higher levels of ambient noise, such as documentation of noise level via the application, and testing capabilities at higher frequencies, such as 6000 and 8000 Hz.

## Conclusions

The accuracy and test–retest reliability of a new component of the NIH Toolbox, the HTT, was assessed using electroacoustic analyses and behavioral validation involving

a clinical sample of children and adults who had varying levels of hearing sensitivity. The HTT is available as part of the NIH Toolbox App via the Apple iPad. The electroacoustic analysis indicated that the HTT met current conformity guidelines for output and linearity for clinical devices that are used to measure hearing. The behavioral validation indicated that the magnitude of differences between the HTT and the gold-standard diagnostic audiological assessment was within current clinical guidelines for children and adults. The sensitivity and specificity of the HTT were high for both children and adults, indicating the results can provide a reliable estimate of the presence and absence of hearing loss and the level of hearing loss within approximately 10 dB. In approximately 5 min, the HTT is able to provide researchers with a validated tool to assess hearing sensitivity across a range of clinical populations and can help to expand the inclusion of people with hearing loss in research in fields that study development, behavior, cognition, and language.

**Acknowledgements** Thanks to the ARMADA study team and developers of the NIH Toolbox applications. We also thank the participants of this study.

**Author note** The authors have no financial relationships relevant to this article to disclose. This study is funded with Federal funds from National Institute on Aging (NIH Grant No. 1U2CAG057441) and the National Institute on Deafness and Other Communication Disorders (R01DC018330, P20GM109023).

**Funding** This study is funded with Federal funds from National Institute on Aging (NIH Grant No. 1U2CAG057441) and the National Institute on Deafness and Other Communication Disorders (R01DC018330, P20GM109023).

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Ethics approval** The study was approved by the Institutional Review Board at Boys Town National Research Hospital (Protocol # 12-05 - XP).

**Consent to participate** Informed consent/assent was obtained from all individual participants included in the study. For children under age 18 years, informed consent was obtained from legal guardians.

## References

- Abu-Ghanem, S., Handzel, O., Ness, L., Ben-Artzi-Blima, M., Fait-Ghelbendorf, K., & Himmelfarb, M. (2016). Smartphone-based audiometric test for screening hearing loss in the elderly. *European Archives of Oto-Rhino-Laryngology*, 273(2), 333–339. <https://doi.org/10.1007/s00405-015-3533-9>
- Amieva, H., Ouvrard, C., Giuliani, C., Meillon, C., Rullier, L., & Dartigues, J. F. (2015). Self-reported hearing loss, hearing aids, and cognitive decline in elderly adults: A 25-year study. *Journal of*

- the American Geriatrics Society*, 63(10), 2099–2104. <https://doi.org/10.1111/jgs.13649>
- Amieva, H., Ouvrard, C., Meillon, C., Rullier, L., & Dartigues, J. F. (2018). Death, depression, disability, and dementia associated with self-reported hearing problems: A 25-year study. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 73(10), 1383–1389. <https://doi.org/10.1093/gerona/glx250>
- American National Standards Institute (ANSI). (1997). *S3.5-1997 R-1997, American National Standards Methods for the calculation of the articulation index*. American National Standards Institute.
- American National Standards Institute/Acoustical Society of America. (1999). *ANSI/ASA S3. 1-1999 (R2018). Maximum permissible ambient noise levels for audiometric test rooms*. American National Standards Institute.
- American National Standards Institute (ANSI). (2018). *S3.6-2018 R-2018, American National Standard Specification for audiometers*. American National Standards Institute.
- ASHA. (2005). *Guidelines for manual pure-tone threshold audiometry*. <https://www.asha.org/policy/g12005-00014/>
- ASHA. (n.d.). *Adult Hearing Screening*. <https://www.asha.org/Practice-Portal/Professional-Issues/Adult-Hearing-Screening>
- Beahan, N., Kei, J., Driscoll, C., Charles, B., & Khan, A. (2012). High-Frequency Pure-Tone Audiometry in Children: A Test–Retest Reliability Study Relative to Ototoxic Criteria. *Ear Hearing*, 33(1), 104–111. <https://doi.org/10.1097/AUD.0b013e318228a77d>
- Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135–160. <https://doi.org/10.1177/096228029900800204>
- Brender, E., Burke, A. E., & Glass, R. M. (2006). Audiometry. *JAMA*, 295(4), 460–460. <https://doi.org/10.1001/jama.295.4.460>
- Bright, T., & Pallawela, D. (2016). Validated Smartphone–Based Apps for Ear and Hearing Assessments: A Review. *JMIR Rehabilitation and Assistive Technologies*, 3(2), e13. <https://doi.org/10.2196/rehab.6074>
- Buss, E., Porter, H. L., Leibold, L. J., Grose, J. H., & Hall, J. W. (2016). Effects of Self–Generated Noise on Estimates of Detection Threshold in Quiet for School–Age Children and Adults. *Ear & Hearing*, 37(6), 650–659. <https://doi.org/10.1097/AUD.0000000000000337>
- Carhart, R., & Jerger, J. F. (1959). Preferred method for clinical determination of pure–tone thresholds. *Journal of Speech and Hearing Disorders*, 24(4), 330–345. <https://doi.org/10.1044/jshd.2404.330>
- Champlin, C., & Letowski, T. (2014). Audiometric calibration: Air conduction. *Seminars in Hearing*, 35, 312–328. <https://doi.org/10.1055/s-0034-1390161>
- Clark, K., Sowers, M., Wallace, R. B., & Anderson, C. (1991). The accuracy of self-reported hearing loss in women aged 60–85 years. *American Journal of Epidemiology*, 134(7), 704–708. <https://doi.org/10.1093/oxfordjournals.aje.a116147>
- Cohn, E. S. (1999). Hearing loss with aging: Presbycusis. *Clinics in Geriatric Medicine*, 15(1), 145–161 viii.
- Corona, A. P., Ferrite, S., Bright, T., & Polack, S. (2020). Validity of hearing screening using hearTest smartphone-based audiometry: Performance evaluation of different response modes. *International Journal of Audiology*, 59(9), 666–673. <https://doi.org/10.1080/14992027.2020.1731767>
- Choi, J. S., Betz, J., Deal, J., Contrera, K. J., Genter, D. J., Chen, D. S., ..., Lin, F. R. (2016). A comparison of self-report and audiometric measures of hearing and their associations with functional outcomes in older adults. *Journal of Aging and Health*, 28(5), 890–910. <https://doi.org/10.1177/0898264315614006>
- Dawes, P., Emsley, R., Cruickshanks, K. J., Moore, D. R., Fortnum, H., Edmondson-Jones, M., ..., Munro, K. J. (2015). Hearing loss and cognition: The role of hearing aids. *Social Isolation and Depression. PLoS One*, 10(3), e0119616. <https://doi.org/10.1371/journal.pone.0119616>
- Eckert, M. A., Matthews, L. J., & Dubno, J. R. (2017). Self-assessed hearing handicap in older adults with poorer-than-predicted speech recognition in noise. *Journal of Speech, Language, and Hearing Research*, 60(1), 251–262. [https://doi.org/10.1044/2016\\_JSLHR-H-16-0011](https://doi.org/10.1044/2016_JSLHR-H-16-0011)
- Feltner, C., Wallace, I. F., Kistler, C. E., Coker-Schwimmer, M., & Jonas, D. E. (2021). Screening for hearing loss in older adults: Updated evidence report and systematic review for the US preventive services task force. *JAMA*, 325(12), 1202–1215. <https://doi.org/10.1001/jama.2020.24855>
- Foulad, A., Bui, P., & Djalilian, H. (2013). Automated audiometry using apple iOS-based application technology. *Otolaryngology–Head and Neck Surgery*, 149(5), 700–706. <https://doi.org/10.1177/0194599813501461>
- Frank, T., Durrant, J. D., & Lovrinic, J. M. (1993). Maximum permissible ambient noise levels for audiometric test rooms. *American Journal of Audiology*, 2(1), 33–37. <https://doi.org/10.1044/1059-0889.0201.33>
- Frank, T. (2000). ANSI update: Maximum permissible ambient noise levels for audiometric test rooms. *American Journal of Audiology*, 9(1), 3–8. [https://doi.org/10.1044/1059-0889\(2000\)003](https://doi.org/10.1044/1059-0889(2000)003)
- Fredriksson, S., Hammar, O., Magnusson, L., Kähäri, K., & Persson Waye, K. (2016). Validating self-reporting of hearing-related symptoms against pure-tone audiometry, otoacoustic emission, and speech audiometry. *International Journal of Audiology*, 55(8), 454–462. <https://doi.org/10.1080/14992027.2016.1177210>
- Füllgrabe, C. (2020). On the possible overestimation of cognitive decline: The impact of age-related hearing loss on cognitive-test performance. *Frontiers in Neuroscience*, 14. <https://doi.org/10.3389/fnins.2020.00454>
- Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., & Wagster, M. V. (2010). Assessment of neurological and behavioural function: The NIH Toolbox. *Lancet Neurology*, 9(2), 138–139. [https://doi.org/10.1016/s1474-4422\(09\)70335-7](https://doi.org/10.1016/s1474-4422(09)70335-7)
- Harrell, R. W. (2002). Pure tone evaluation. In J. Katz (Ed.), *Handbook of clinical audiology* (5th ed., pp. 71–87). Lippincott Williams & Wilkins.
- Hughson, W., & Westlake, H. D. (1944). Manual for program outline for rehabilitation of aural casualties both military and civilian. *Transactions - American Academy of Ophthalmology and Otolaryngology*, 48, 1–15.
- Trace, A. L., Sharma, R. K., Reed, N. S., & Golub, J. S. (2021). Smartphone-based applications to detect hearing loss: A review of current technology. *Journal of the American Geriatrics Society*, 69(2), 307–316. <https://doi.org/10.1111/jgs.16985>
- Jerger, J. (2018). The evolution of the audiometric pure-tone technique. *The Hearing Journal*, 25(9), 12–18.
- Kamil, R. J., Genter, D. J., & Lin, F. R. (2015). Factors associated with the accuracy of subjective assessments of hearing impairment. *Ear & Hearing*, 36(1), 164–167. <https://doi.org/10.1097/aud.0000000000000075>
- Kiely, K. M., Gopinath, B., Mitchell, P., Browning, C. J., & Anstey, K. J. (2012). Evaluating a dichotomized measure of self-reported hearing loss against gold standard audiometry: Prevalence estimates and age bias in a pooled National Data set. *Journal of Aging and Health*, 24(3), 439–458. <https://doi.org/10.1177/0898264311425088>
- Kronenberger, W. G., & Pisoni, D. B. (2020). Why are children with cochlear implants at risk for executive functioning delays? Language only or something more? In M. Marshark & H. Knoors (Eds.), *The oxford handbook of deaf studies in learning and cognition* (pp. 248–267). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190054045.013.17>
- Lawrence, B. J., Jayakody, D. M. P., Bennett, R. J., Eikelboom, R. H., Gasson, N., & Friedland, P. L. (2020). Hearing loss and

- depression in older adults: A systematic review and meta-analysis. *Gerontologist*, 60(3), e137–e154. <https://doi.org/10.1093/geront/gnz009>
- Lin, F. R., Metter, E. J., O'Brien, R. J., Resnick, S. M., Zonderman, A. B., & Ferrucci, L. (2011). Hearing loss and incident dementia. *Archives of Neurology*, 68(2), 214–220. <https://doi.org/10.1001/archneurol.2010.362>
- Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., ... Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet commission. *The Lancet*, 396(10248), 413–446. [https://doi.org/10.1016/S0140-6736\(20\)30367-6](https://doi.org/10.1016/S0140-6736(20)30367-6)
- Louw, C., Swanepoel, D. W., & Eikelboom, R. H. (2018). Self-reported hearing loss and pure tone audiometry for screening in primary health care clinics. *Journal of Primary Care & Community Health*, 9, 2150132718803156. <https://doi.org/10.1177/2150132718803156>
- Maharani, A., Pendleton, N., & Leroi, I. (2019). Hearing impairment, loneliness, social isolation, and cognitive function: Longitudinal analysis using English longitudinal study on ageing. *The American Journal of Geriatric Psychiatry*, 27(12), 1348–1356. <https://doi.org/10.1016/j.jagp.2019.07.010>
- Mahomed, F., Swanepoel, D. W., Eikelboom, R. H., & Soer, M. (2013). Validity of automated threshold audiometry: A systematic review and meta-analysis. *Ear & Hearing*, 34(6), 745–752. <https://doi.org/10.1097/01.aud.0000436255.53747.a4>
- Martin, F. N., & Clark, J. G. (2019). *Introduction to audiology* (13th Edition ed.). Pearson.
- McCreery, R. W., & Walker, E. A. (2021). Variation in auditory experience affects language and executive function skills in children who are hard of hearing. *Ear & Hearing*. <https://doi.org/10.1097/aud.0000000000001098>
- McCreery, R. W., Walker, E. A., Spratford, M., Bentler, R., Holte, L., Roush, P., ..., Moeller, M. P. (2015). Longitudinal predictors of aided speech audibility in infants and children. *Ear & Hearing*, 36, 24S–37S. <https://doi.org/10.1097/aud.0000000000000211>
- McCreery, R. W., Walker, E. A., Stiles, D. J., Spratford, M., Oleson, J. J., & Lewis, D. E. (2020). Audibility-based hearing aid fitting criteria for children with mild bilateral hearing loss. *Language Speech and Hearing Services in Schools*, 51(1), 55–67. [https://doi.org/10.1044/2019\\_lshss-ochl-19-0021](https://doi.org/10.1044/2019_lshss-ochl-19-0021)
- Moore, D. R., Zobay, O., & Ferguson, M. A. (2020). Minimal and mild hearing loss in children: Association with auditory perception, cognition, and communication problems. *Ear & Hearing*, 41(4), 720–732. <https://doi.org/10.1097/aud.00000000000000802>
- National Institutes of Health & Northwestern University. (2021). NIH Toolbox (Version 1.23.4940) [Mobile app]. <https://apps.apple.com/us/app/nih-toolbox/id1002228307>
- Nondahl, D. M., Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, R., & Klein, B. E. (1998). Accuracy of self-reported hearing loss. *Audiology*, 37(5), 295–301. <https://doi.org/10.3109/00206099809072983>
- Peer, S., & Fagan, J. J. (2015). Hearing loss in the developing world: Evaluating the iPhone mobile device as a screening tool. *South African Medical Journal*, 105(1), 35–39. <https://doi.org/10.7196/samj.8338>
- Pittman, A. L., Lewis, D. E., Hoover, B. M., & Stelmachowicz, P. G. (2005). Rapid word-learning in normal-hearing and hearing-impaired children: Effects of age, receptive vocabulary, and high-frequency amplification. *Ear & Hearing*, 26(6), 619–629. <https://doi.org/10.1097/01.aud.0000189921.34322.68>
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Reilly, J., Troiani, V., Grossman, M., & Wingfield, A. (2007). An introduction to hearing loss and screening procedures for behavioral research. *Behavior Research Methods*, 39(3), 667–672. <https://doi.org/10.3758/BF03193038>
- Roeser, R. J. (2013). *Roeser's audiology desk reference* (2nd ed.). Thieme.
- Saliba, J., Al-Reefi, M., Carriere, J. S., Verma, N., Provencal, C., & Rappaport, J. M. (2017). Accuracy of Mobile-based audiometry in the evaluation of hearing loss in quiet and Noisy environments. *Otolaryngology–Head and Neck Surgery*, 156(4), 706–711. <https://doi.org/10.1177/0194599816683663>
- Salonen, J., Johansson, R., Karjalainen, S., Vahlberg, T., & Isoaho, R. (2011). Relationship between self-reported hearing and measured hearing impairment in an elderly population in Finland. *International Journal of Audiology*, 50(5), 297–302. <https://doi.org/10.3109/14992027.2010.549517>
- Sandström, J., Swanepoel, D., Laurent, C., Umefjord, G., & Lundberg, T. (2020). Accuracy and reliability of smartphone self-test audiometry in community clinics in low income settings: A comparative study. *Annals of Otolaryngology & Rhinology*, 129(6), 578–584. <https://doi.org/10.1177/0003489420902162>
- Schnohr, C., Frid, K. C., Jensen, R. G., Homøe, P., & Niclasen, B. V. L. (2019). Self-reported hearing impairment among Greenlandic adolescents: Item development and findings from the health behaviour in school-aged children study 2018. *International Journal of Pediatric Otorhinolaryngology*, 126, 109629. <https://doi.org/10.1016/j.ijporl.2019.109629>
- Stiles, D. J., McGregor, K. K., & Bentler, R. A. (2013). Wordlikeness and word learning in children with hearing loss. *International Journal of Language & Communication Disorders*, 48(2), 200–206. <https://doi.org/10.1111/j.1460-6984.2012.00199.x>
- Swanepoel, D. W., Myburgh, H. C., Howe, D. M., Mahomed, F., & Eikelboom, R. H. (2014). Smartphone hearing screening with integrated quality control and data management. *International Journal of Audiology*, 53(12), 841–849. <https://doi.org/10.3109/14992027.2014.920965>
- Szudek, J., Ostevik, A., Dziegielewska, P., Robinson-Anagor, J., Goma, N., Hodgetts, B., & Ho, A. (2012). Can Uhear me now? Validation of an iPod-based hearing loss screening test. *Journal of Otolaryngology - Head & Neck Surgery*, 41(Suppl 1), S78–S84.
- Thompson, G. P., Sladen, D. P., Borst, B. J., & Still, O. L. (2015). Accuracy of a tablet audiometer for measuring behavioral hearing thresholds in a clinical population. *Otolaryngology–Head and Neck Surgery*, 153(5), 838–842. <https://doi.org/10.1177/0194599815593737>
- Thomson, R. S., Auduong, P., Miller, A. T., & Gurgel, R. K. (2017). Hearing loss as a risk factor for dementia: A systematic review. *Laryngoscope Investigative Otolaryngology*, 2(2), 69–79. <https://doi.org/10.1002/lio2.65>
- Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., & Moeller, M. P. (2015). Language outcomes in young children with mild to severe hearing loss. *Ear & Hearing*, 36, 76S–91S. <https://doi.org/10.1097/AUD.0000000000000219>
- Tsimpida, D., Kontopantelis, E., Ashcroft, D., & Panagioti, M. (2020). Comparison of self-reported measures of hearing with an objective audiometric measure in adults in the English longitudinal study of ageing. *JAMA Network Open*, 3(8), e2015009–e2015009. <https://doi.org/10.1001/jamanetworkopen.2020.15009>
- Vaccaro, R., Zaccaria, D., Colombo, M., Abbondanza, S., & Guaita, A. (2019). Adverse effect of self-reported hearing disability in elderly Italians: Results from the InveCe.Ab study. *Maturitas*, 121, 35–40. <https://doi.org/10.1016/j.maturitas.2018.12.009>
- Vassilaki, M., Aakre, J. A., Knopman, D. S., Kremers, W. K., Mielke, M. M., Geda, Y. E., ..., Petersen, R. C. (2019). Informant-based hearing difficulties and the risk for mild cognitive impairment

- and dementia. *Age and Ageing*, 48(6), 888–894. <https://doi.org/10.1093/ageing/afz099>
- Weintraub, S., Karpouzian-Rogers, T., Peipert, J. D., Nowinski, C., Slotkin, J., Wortman, K., ..., Gershon, R. (2021). ARMADA: Assessing reliable measurement in Alzheimer's disease and cognitive aging project methods. *Alzheimer's & Dementia*, 1–12. <https://doi.org/10.1002/alz.12497>.
- World Health Organization. (1991). *Report of the informal working group on prevention of deafness and hearing impairment Programme planning*. Geneva: World Health Organization <https://apps.who.int/iris/handle/10665/58839>.
- World Health Organization. (2001). International classification of functioning, disability and health (ICF). <https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health>.
- Yeung, J. C., Heley, S., Beaugard, Y., Champagne, S., & Bromwich, M. A. (2015). Self-administered hearing loss screening using an interactive, tablet play audiometer with ear bud headphones. *International Journal of Pediatric Otorhinolaryngology*, 79(8), 1248–1252. <https://doi.org/10.1016/j.ijporl.2015.05.021>
- Yousuf Hussein, S., Wet Swanepoel, D., Biagio de Jager, L., Myburgh, H. C., Eikelboom, R. H., & Hugo, J. (2016). Smartphone hearing screening in mHealth assisted community-based primary care. *Journal of Telemedicine and Telecare*, 22(7), 405–412. <https://doi.org/10.1177/1357633x15610721>
- Zecker, S. G., Hoffman, H. J., Frisina, R., Dubno, J. R., Dhar, S., Wallhagen, M., ..., Wilson, R. H. (2013). Audition assessment using the NIH Toolbox. *Neurology*, 80(11 Suppl 3), S45–S48. <https://doi.org/10.1212/WNL.0b013e3182872dd2>
- Zekveld, A. A., George, E. L., Houtgast, T., & Kramer, S. E. (2013). Cognitive abilities relate to self-reported hearing disability. *Journal of Speech, Language, and Hearing Research*, 56(5), 1364–1372. [https://doi.org/10.1044/1092-4388\(2013\)12-0268](https://doi.org/10.1044/1092-4388(2013)12-0268)

**Open practices statement:** None of the experiments were preregistered. The deidentified data and materials for all experiments are available at [https://osf.io/dtmv9/?view\\_only=90c3b1a1add7420d988087f276c517ed](https://osf.io/dtmv9/?view_only=90c3b1a1add7420d988087f276c517ed)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.