Development and Validation of the Effectiveness of Auditory Rehabilitation Scale

Bevan Yueh, MD, MPH; Jennifer A. McDowell, MS; Margaret Collins, MS, CCC-A; Pamela E. Souza, PhD, CCC-A; Carl F. Loovis, PhD, CCC-A; Richard A. Deyo, MD, MPH

Objective: To develop a new scale of hearing-related function and quality of life in patients with hearing aids that addresses overlooked concerns, such as hearing-aid comfort, convenience, and cosmetic appearance, that may influence hearing-aid adherence while maintaining brevity and sensitivity to clinical change.

Design: Prospective, multicenter instrument validation.

Setting: Four diverse sites in Washington State, including 2 private practices, 1 university setting, and 1 Veterans Affairs hospital.

Patients: Seventy-eight patients with hearing aids.

Interventions: We created 2 modules in the Effectiveness of Auditory Rehabilitation (EAR) scale. The first module (Inner EAR) covers intrinsic hearing issues such as hearing in quiet and hearing in noise and is administered both before and after treatment. The second module (Outer EAR) covers extrinsic (hearing-aid related) issues such as comfort, appearance, and convenience and is administered after hearing-aid fitting.

Main Outcome Measures: Both scales were developed and validated in 3 stages. Stage 1 used a qualitative approach from multiple data sources to develop preliminary instruments. Stage 2 used approaches from classic test theory to reduce the number of items and psychometrically validate the instruments. Stage 3 examined the responsiveness or sensitivity to clinical change.

Results: A 10-item Inner EAR module and a 10-item Outer EAR module were created and validated. Internal consistency of individual domains (Cronbach α=0.85 and 0.72, respectively) and test-retest reliability (intraclass correlation coefficients=0.76 and 0.81, respectively) were excellent. Evidence of construct validity included concurrent validity with other hearing scales and global visual analog scales, discriminant validity with dizziness handicap, correlation with hearing-aid adherence, and confirmatory factor analyses. Both scales had strong evidence of responsiveness (sensitivity to change), with higher effect sizes and Guyatt responsiveness statistics than the 2 widely used hearing scales in this study. The scales took an average of 5 minutes to complete.

Conclusions: The EAR scale is a valid and reliable measure of the effectiveness of amplification in the treatment of sensorineural hearing loss. It addresses the range of issues that are of importance to hearing-aid patients. The scales have excellent psychometric properties, are more responsive than several widely used hearing scales, and are minimally burdensome for patients to complete. The EAR may be a valuable outcome measure in future studies of both existing hearing aids and newer hearing-aid technologies, such as bone-anchored aids or middle ear implants.


SENSORINEURAL HEARING LOSS is one of the most common chronic disabilities in the United States. More than 25% of the population older than 65 years is hearing impaired. The diminished ability to hear and communicate is frustrating in and of itself, but the strong association of hearing loss with functional decline and psychiatric morbidity adds further to the burden on the hearing-impaired population. Fortunately, treatment of sensorineural hearing loss with hearing aids reverses many of its negative effects, leading to better overall quality of life.

An impressive number of outcome measures have been developed for patients with sensorineural hearing impairment and those who wear hearing aids, and many...
of these scales have been widely applied. Although the focus on hearing-related effects has successfully differentiated the quality of life of hearing-aid users from non-users,11-14 these scales may not be sensitive to differences among hearing aids.9,30 Furthermore, data from a prior randomized trial11 suggest that existing scales may not address the issues that are paramount to patients who stopped using hearing aids. For example, comfort, convenience, and cosmetic appearance are typically not considered, even though they may affect adherence as much as or more than purely acoustic concerns. All hearing specialists are familiar with patients who test well in a sound booth and respond well to hearing questionnaires but refuse to wear the aids because it makes them “look old.”

The purpose of this project was to develop a scale to assess hearing-related function and quality of life in patients with sensorineural hearing loss. We sought to develop a new scale that addressed all aspects of treatment with hearing aids, including comfort, convenience, and cosmetics. In addition, we sought to create a scale that had not only strong psychometric properties of reliability and validity but also excellent responsiveness while causing minimal patient burden. We believe that the resulting scale may ultimately be useful for comparing emerging hearing-aid technologies, such as bone-anchored aids, middle ear implants, and disposable hearing aids.

### METHODS

The Effectiveness of Auditory Rehabilitation (EAR) scale was created in 3 phases: (1) questionnaire development, (2) instrument validation, and (3) responsiveness testing. Patients in phase 1 consisted of those participating in a randomized trial of hearing amplification options at a Department of Veterans Affairs (VA) hospital13 and a convenience sample of patients at the University of Washington. Phase 2 and 3 patients were consecutively enrolled from July 2000 to November 2001 from 4 diverse sites of practice in Washington State to improve generalizability. The sites were VA Puget Sound (Seattle), the University of Washington (Seattle), a private otolaryngology and audiology practice in eastern Washington (Spokane), and a hearing-aid dispensing practice (Tacoma).

All patients who presented for hearing-aid evaluations were considered good if in the absence of interim treatment. Internal consistency is considered good if α approximates 0.70 but does not exceed 0.90, because values higher than 0.90 imply the presence of redundant items.13 Items that did not substantially improve the internal consistency of a particular domain were eliminated. Test-retest reliability was measured with the ICC, which is more rigorous than the Pearson correlation coefficient (r) because it considers not just the strength of the correlation but also systematic variations.31

The 3 forms of validity that we sought were content, criterion, and construct. Content validity was established with the rigorous approach to item development in phase 1. Criterion validity, which tests scale performance compared with a gold standard, is difficult to establish when evaluating quality-of-life scales.

Construct validity is present if the scale behaves according to hypothesized relationships. We hypothesized that the Inner EAR would correlate (r ≥ 0.35) with a summary rating scale of overall hearing function; the Outer EAR would correlate (r ≥ 0.35) with summary rating scales of hearing improvement; the Inner EAR would correlate (r ≥ 0.35) with hearing handicap (measured with the Hearing Handicap Inventory for the Elderly [HHIE]13-15), since both measures reflect hearing

### PHASE 1: QUESTIONNAIRE DEVELOPMENT

We adopted a qualitative approach to identify domains of greatest relevance to stakeholders, including patients, audiologists, and otolaryngologists. We analyzed items from existing scales, data from patients’ open-ended questionnaires about their hearing-aid experiences, diaries of hearing-aid use, and interviews to create a taxonomy of factors that affect the quality of life of hearing-aid users.13 The 4 study authors who are also hearing-loss professionals (B.Y., P.E.S., M.C., and C.F.L.) then met weekly for a month to identify other factors of clinical interest to ensure “theme saturation,” at which point no additional factors emerged. We created multiple questions in each domain, and nearly 100 questions were administered to the convenience cohort. One scale was developed to address intrinsic hearing issues that could be assessed with or without a hearing aid (Inner EAR), and a second scale was created to address only the extrinsic issues associated with use of a hearing aid (Outer EAR). Low response rates, inconsistent responses, skewed distributions, and ceiling and floor effects were used to identify and eliminate suboptimal questions. We adhered to the fundamental principle that the scales should have maximum clinical utility, and we therefore sought to keep the modules as brief as possible to minimize patient burden. This iterative testing process was also used to refine the response options for each question.

### PHASE 2: INSTRUMENT VALIDATION

We adhered to the tenets of classical test theory to validate both EAR modules. Reliability was established by ensuring both internal consistency (Cronbach α) and test-retest reliability (intraclass correlation coefficient [ICC]) at 2 weeks (Table 1) in the absence of interim treatment. Internal consistency is considered good if α approximates 0.70 but does not exceed 0.90, because values higher than 0.90 imply the presence of redundant items.13 Items that did not substantially improve the internal consistency of a particular domain were eliminated. Test-retest reliability was measured with the ICC, which is more rigorous than the Pearson correlation coefficient (r) because it considers not just the strength of the correlation but also systematic variations.31

The 3 forms of validity that we sought were content, criterion, and construct. Content validity was established with the rigorous approach to item development in phase 1. Criterion validity, which tests scale performance compared with a gold standard, is difficult to establish when evaluating quality-of-life scales.

Construct validity is present if the scale behaves according to hypothesized relationships. We hypothesized that the Inner EAR would correlate (r ≥ 0.35) with a summary rating scale of overall hearing function; the Outer EAR would correlate (r ≥ 0.35) with summary rating scales of hearing improvement; the Inner EAR would correlate (r ≥ 0.35) with hearing handicap (measured with the Hearing Handicap Inventory for the Elderly [HHIE]13-15), since both measures reflect hearing...
function; the Outer EAR would correlate ($r \geq 0.35$) with hearing-aid adherence, since unsatisfactory aids should lead to nonadherence; and both scales would correlate ($r \geq 0.35$) with communication benefit (measured with the Abbreviated Profile of Hearing Aid Benefit [APHAB][16,39]). Because strong correlations are anticipated, this type of validity is termed concurrent validity. We also hypothesized that EAR scores would not correlate with nonhearing but potentially correlated constructs such as dizziness (discriminant validity). We used the Dizziness Handicap Inventory (DHI)35 to assess dizziness. Therefore, the EAR modules, HHI, APHAB, DHI, and adherence data were collected at both baseline and final visits (Table 1). Finally, we used factor analyses to confirm the presence of hypothesized domains and subdomains. We used a principal components technique with an orthogonal varimax rotation, retaining factors with eigenvalues greater than 1.0.

**PHASE 3: RESPONSIVENESS TESTING**

Responsiveness, or sensitivity to clinical change, is important because scales must be able to detect clinically meaningful changes in the health state. Because performance with hearing-aid use may improve during the first few months of use,[36,37] we assessed outcomes 3 months after hearing aids were received (Table 1). Changes in scores from baseline to 3 months were calculated. Responsiveness was assessed with the effect size (mean change divided by standard deviation of the Inner EAR score at baseline) and the Guyatt responsiveness statistic (mean change divided by standard deviation of change).[38,39]

**SAMPLE SIZE**

We based sample size estimates on collecting adequate data for the Outer EAR. (Because the Inner EAR applies to both new and experienced users and the Outer EAR is relevant before index treatment only for experienced users, achieving adequate statistical power for the Outer EAR ensured adequate power for the Inner EAR analyses.) With a 1-tailed $\alpha$ of .05 and a $\beta$ of .20 (80% power), 44 experienced users were needed to detect an $r \geq 0.35$ to achieve construct (concurrent) validity.[39] A 1-tailed test was used because we hypothesized that the correlation would be at least 0.35. The power to detect a statistically significant result was greater for the Inner EAR, since the whole population (both new and experienced users) was used.

**DATA MANAGEMENT**

Data were entered via double-entry verification for accuracy and maintained in a password-protected local area network at VA Puget Sound. Analyses were performed with standard statistical software (STATA/SE 8.2 for Windows; Stata Corp, College Station, Tex). The study was approved by institutional review boards at the VA Puget Sound and University of Washington.

**RESULTS**

**PHASE 1: QUESTIONNAIRE DEVELOPMENT**

Qualitative analyses of interview, diary, and opened questionnaire data collected in a prior hearing-aid trial[13] revealed that existing scales did not address numerous issues of importance to hearing-aid users. For example, in exit interviews of nonadherent patients, 5 common explanations were given. First came “just forgetting,” but the next 3 most common reasons were physical discomfort from wearing the aid, difficulties with manipulation, and inconvenience of the device. Of note, the complaint that hearing did not improve with the aids—the primary focus of existing scales—was only the fifth most common reason for nonadherence.

We used a clinimetric approach[40] to organize these data into categories based on similar themes to create a taxonomy of categories (domains) that affect the quality of life of hearing-aid users. Two general categories (parts 1 and 2) were identified. Part 1 included issues intrinsic to hearing loss, including speech understanding and the socioemotional impact of hearing loss. Part 2 included extrinsic issues associated with an amplification device, including amplification acoustics, convenience, comfort, appearance and self-esteem, and maintenance and reliability of the aid. A review of existing hearing-related outcome measures found that although part 1 issues were usually addressed, part 2 issues were rarely covered. These findings pointed to the need for an instrument that considered the convenience, comfort, appearance, reliability, and technological issues that patients believe are important, especially since dissatisfaction with these factors may result in nonadherence.

We noted that issues in part 1 of the taxonomy applied to both patients with and without hearing aids but part 2 issues were relevant only to hearing-aid users. Therefore, 2 modules were created. The Inner EAR would address intrinsic issues related to hearing, such as speech understanding and socioemotional effects of hearing impairment. The Outer EAR would cover extrinsic factors, such as comfort, convenience, and cosmetic appearance, and would be used only for hearing-aid users. The final versions of the Inner and Outer EAR each had 10 individual questions, along with 1 and 2 global scales for each measure, respectively (Table 2).

**PHASE 2: INSTRUMENT VALIDATION**

Of 105 patients who were offered enrollment, 78 subjects completed the study; 18 refused participation, and 9 enrolled but did not complete their final visit. Mean hearing thresholds are shown in the Figure. Enrollment was proportional to clinical volume at each of the 4 sites. Of the patients who completed the study, 38 came from the VA, 7 from the University of Washington, 23 from the private practice in Spokane, and 10 from the hearing-aid dispensing practice in Tacoma (Table 3). Of these 78 participants, 46 (59%) were experienced users. The mean time required to complete each EAR module was 5 minutes.

**SCORING OF EAR MODULES**

Composite scores of each module consisted of the mean score of the 10 individual items (not including the global scales). Individual items are scored on a 0 to 100 scale; for instance, if there are 5 responses, the responses are given scores of 0, 25, 50, 75, and 100, with 100 representing the best function or rating. Thus, a score of 100 is the best possible result; a score of 0 is the worst possible result.

Internal consistency was measured by the Cronbach $\alpha$, which was 0.85 for the Inner EAR and 0.72 for the Outer...
Table 2. Items on the Inner and Outer Effectiveness of Auditory Rehabilitation (EAR) Scale

<table>
<thead>
<tr>
<th>Inner EAR</th>
<th>How would you rate your ability to understand the following?</th>
<th>What family and close friends are saying</th>
<th>Speech in a quiet room</th>
<th>What is said by those at your table in a crowded restaurant</th>
<th>What you want to hear and filter out unwanted noises</th>
<th>Telephone conversations</th>
<th>How would you rate your ability to hear the following?</th>
<th>Your ability to hear in different listening situations</th>
<th>Soft household sounds (car turn signal or clock ticking)</th>
<th>How would you rate your mood based on your ability to hear?</th>
<th>How bothered are you by the following?</th>
<th>Having to ask people to repeat things</th>
<th>Having to restrict your activities because of your hearing</th>
<th>How would you rate your ability to hear? (global)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer EAR</td>
<td>How would you rate the following?</td>
<td>Quality of your own voice</td>
<td>How natural your hearing device sounds</td>
<td>The physical comfort of your device after wearing it for a while</td>
<td>The convenience of your hearing device</td>
<td>The dependability of your hearing device</td>
<td>How bothered have you been by the following?</td>
<td>Loud noises such as a door slamming or a pan dropping</td>
<td>Wind noise</td>
<td>How you look with your hearing device</td>
<td>Having to change its batteries</td>
<td>Your hearing device making a squealing noise</td>
<td>How would you rate the improvement in your hearing? (global)</td>
<td>Was your hearing device worth the investment? (global)</td>
</tr>
</tbody>
</table>

Table 3. Demographic Characteristics of Patients by Site

<table>
<thead>
<tr>
<th>Clinic</th>
<th>Total No. of Patients</th>
<th>No. of Experienced Users</th>
<th>Mean Age, y</th>
<th>Male, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Veterans Affairs Puget Sound</td>
<td>38</td>
<td>23</td>
<td>74.0</td>
<td>97</td>
</tr>
<tr>
<td>University of Washington</td>
<td>7</td>
<td>5</td>
<td>81.1</td>
<td>57</td>
</tr>
<tr>
<td>Private otolaryngology practice</td>
<td>23</td>
<td>9</td>
<td>68.8</td>
<td>74</td>
</tr>
<tr>
<td>Hearing-aid dispensing practice</td>
<td>10</td>
<td>9</td>
<td>66.1</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>78</td>
<td>46</td>
<td>72.1</td>
<td>81</td>
</tr>
</tbody>
</table>

Figure. Mean hearing thresholds for the right and left ears. Long dashed lines show ±1 SD for the right ear; short dashed lines show ±1 SD for the left ear. HL indicates hearing level.

EAR. Both scores were above 0.70 but not above 0.90, which would suggest the presence of redundant questions. The internal consistency of individual domains for the Inner EAR module also ranged from 0.70 to 0.85. Test-retest reliability was also high. Of the 78 subjects, 60 repeated the Inner EAR before receiving their hearing aid (ICC of 0.76; mean interval between administrations, 11.4 days). Of the 46 experienced patients who completed an Outer EAR at baseline, 32 completed a retest (ICC of 0.81).

Construct validity was first evaluated by comparing composite scores with the global scales on each module using the Pearson correlation coefficient (r). The Inner EAR composite score was compared with the global question, “How would you rate your ability to hear?” (r = 0.60 at baseline and 0.70 at 3 months). The composite Outer EAR score’s correlation coefficients with its 2 global questions, “How would you rate the improvement in your hearing?” and “Was your hearing device worth the investment?” were r = 0.42 and r = 0.47 at baseline and r = 0.74 and r = 0.71 at 3 months, respectively.

Construct validity was also seen with a variety of other hypothesized relationships. High correlations were found between the Inner EAR and the HHIE (r = 0.72 at baseline and 0.54 at 3 months). High correlations were found between both EAR modules and the APHAB at baseline and 3 months (r = 0.47-0.61). Correlations greater than the hypothesized values of 0.30 were found at 3 months between the Outer EAR and adherence with days per week of use (r = 0.32) and hours per week of use (r = 0.46). Low correlations were found with dizziness (r = 0.04-0.27).

Factor analysis was used to confirm that the items loaded into 2 major domains (Inner and Outer EAR). We also sought to identify subdomains within each module. With a principal components analysis using an orthogonal varimax rotation, we identified 2 factors with an eigenvalue greater than 1.0 in the Inner EAR, which confirmed the presence of speech understanding and socioemotional domains, as expected. However, because of the number of hypothesized factors in the Outer EAR and the relatively small number of patients, we were unable to identify consistent subdomains within the Outer EAR.

**PHASE 3: RESPONSIVENESS TESTING**

Both the Inner and Outer EAR showed evidence of high sensitivity to change (responsiveness), as measured by both the effect size and the Guyatt responsiveness statistics (Table 4). The effect size for the Inner EAR was very high (>2.1). The effect size for the Outer EAR was expected to be lower, since only patients who already had hearing aids filled out a baseline Outer EAR questionnaire. Nonetheless, the effect size was still higher than 1.1, in contrast to the values for the HHIE and APHAB (effect sizes < 0.9). The
loss and hearing aids. We do not anticipate that the psy-
founding from patients with different types of hearing
This extended our enrollment period but reduced con-
enrollment to patients who received bilateral hearing aids.
elderly, but this is less troubling because hearing aids are
sponses or between men and women. The cohort is el-
elastic differences between veteran and nonveteran re-
recruited from the VA, which partially explains the pre-
enced hearing-aid users.
The SADL also contains items about cost and the
fitting. The SADL also contains items about cost and the
item SADL does address some of these issues but, like
dress comfort, convenience, or cosmetic issues. The 15-
function, and hearing-aid adherence. It is highly sensi-
tive to clinical change.
We believe that the EAR offers several unique strengths.
First, it addresses comfort, convenience, and cosmetic is-
ues that are overlooked by existing scales and that may
lead to hearing-aid nonadherence. Second, the modular de-
sign allows administration of appropriate questions both
before (Inner EAR only for naive hearing-aid users and both
modules for experienced users) and after treatment (both
Inner and Outer EAR). Third, the scales have larger effect
sizes and therefore are more responsive to change than ex-
wisting, widely used scales such as the HHIE and APHAB.
Finally, the brevity of the scale minimizes patient burden.
Some excellent scales have been developed to study con-
ductive41,42 and sensorineural17-20 hearing loss. For
hearing-aid patients, several scales in particular, including
the HHIE,15,33 APHAB,16,34 Satisfaction with Amplific-
in Daily Life (SADL),28 and International Outcome
Inventory for Hearing Aids (IOI-HA), 27 have
received widespread attention by the hearing-research
community. Each of these scales is reliable, valid, and
widely used. Each scale has its own advantages and dis-
advantages. The HHIE, APHAB, and IOI-HA do not ad-
ress comfort, convenience, or cosmetic issues. The 15-
item SADL does address some of these issues but, like
the IOI-HA, can be administered only after hearing-aid
fitting. The SADL also contains items about cost and the
perceived competency of the hearing-loss professionals
that may be less pertinent to evaluation of the hearing-
aid technology itself. Furthermore, little is known about
the SADL’s responsiveness to clinical change in experi-
enced hearing-aid users.

Nearly 50% (38 of 78) of the patients in this study were
recruited from the VA, which partially explains the pre-
dominance of elderly male patients in this study. We made
every effort to recruit from 3 nonveteran populations to
improve the generalizability of our findings. However,
the large volume of patients seen at VA Puget Sound com-
pared with the other centers led to greater enrollment of
veterans. Nonetheless, a substantial number of patients
were nonveterans, and we saw no evidence of system-
atic differences between veteran and nonveteran re-
sponses or between men and women. The cohort is el-
derly, but this is less troubling because hearing aids are
typically used by elderly patients. We also restricted our
enrollment to patients who received bilateral hearing aids.
This extended our enrollment period but reduced con-
 founding from patients with different types of hearing
loss and hearing aids. We do not anticipate that the psy-
chometric properties of the EAR will vary systematically in
patients using only 1 hearing aid, but further study of
such patients is needed to verify this assumption.

The EAR scale is a reliable, valid, and responsive dis-
ease-specific measure of the effectiveness of auditory re-
habilitation. We now seek to establish population-
based normative values for both modules and to identify
cohesive subdomains of the Outer EAR. We encourage
the use of these modules to compare different modalit-
tics of hearing amplification, including comparisons of
emerging technologies such as middle ear implants and
bone-anchored hearing aids with conventional hearing
amplification. Since disease-specific and generic mea-
sures provide complementary information about therapy,
we believe that the use of the EAR in conjunction with
generic outcome measures can lead to important in-
sights in future hearing amplification trials.

Submitted for Publication: March 14, 2005; accepted May
25, 2005.

Author Affiliations: Health Services Research and De-
velopment Service (Dr Yueh and Ms McDowell and
Collins), Surgery and Perioperative Care Service
(Dr Yueh), and Rehabilitation Care Service (Dr Loovis),
VA Puget Sound Health Care System, and Department
of Otolaryngology/Head and Neck Surgery (Dr Yueh),
Department of Speech and Hearing Sciences (Ms Collins
and Dr Souza), Department of Internal Medicine (Dr Deyo),
Center for Cost and Outcomes Research (DrS Yueh and
Deyo), and Department of Health Services (Drs Yueh and
Deyo), University of Washington, Seattle.

Correspondence: Bevan Yueh, MD, MPH, Surgery Ser-
vice (112OTO), VA Puget Sound Health Care System,
1660 S Columbian Way, Seattle, WA 98108 (byueh@u.
washington.edu).

Financial Disclosure: None.

Funding/Support: Dr Yueh is supported by career
development award CD-98318 from the Department of
Veterans Affairs, Veterans Health Administration, Health
Services Research and Development Service, Washing-
ton, DC. This work was supported by grant R-03-AG18150
from the National Institute on Aging, National Insti-
tutes of Health, Bethesda, Md.

Disclaimer: The views expressed in this article are those
of the authors and do not necessarily represent the views
of the Department of Veterans Affairs.

Acknowledgment: We are grateful to the VA Puget Sound
Audiology Clinic (Rebecca Coombs, CCC-A, Billie Garber,
REFERENCES


**Correction**

**Error in Title.** In the Original Article by Yueh et al titled “Development and Validation of the Effectiveness of the Auditory Rehabilitation Scale,” published in the October issue of the ARCHIVES (2005;131:851-856), an error occurred in the title. The correct title is “Development and Validation of the Effectiveness of Auditory Rehabilitation Scale.” This correction was made previously to online versions of this article.