Objective and Subjective Hearing Aid Assessment Outcomes

Lisa Lucks Mendel
The University of Memphis, Memphis, TN

Purpose: To determine whether specific sentence recognition assessments were sensitive enough to serve as objective outcome measurements that document subjective improvements in speech understanding with hearing aids.

Method: The Revised Speech Perception in Noise test (R-SPIN; R. C. Bilger, J. M. Nuetzel, W. M. Rabinowitz, & C. Rzeczkowski, 1984), the Hearing in Noise Test (HINT; M. Nilsson, S. D. Soli, & J. A. Sullivan, 1994), and the Quick Speech-in-Noise test (QuickSIN; Etymotic Research, 2001; M. C. Killion, P. A. Niquette, G. I. Gudmundsen, L. J. Revit, & S. Banerjee, 2004) were administered to 21 hearing aid users to determine whether the tests could adequately document improvements in speech understanding with hearing aids compared with the research participants' self-assessments of their own performance. Comparisons were made between unaided and aided performance on these sentence tests and on the Hearing Aid Performance Inventory (HAPI; B. E. Walden, M. Demorest, & E. Hepler, 1984).

Results: The R-SPIN, the HINT Quiet threshold, and the QuickSIN signal-to-noise ratio (SNR) loss were the most sensitive of the sentence recognition tests to objectively assess improvements in speech perception performance with hearing aids. Comparisons among the subjective and objective outcome measures documented that HAPI ratings improved as performance on the R-SPIN, the HINT Quiet threshold, and the QuickSIN SNR loss improved.

Conclusions: Objective documentation of subjective impressions is essential for determining the efficacy of treatment outcomes in hearing aid fitting. The findings reported here more clearly define the relationship between objective and subjective outcome measures in an attempt to better define true hearing aid benefit.

Key Words: outcomes, hearing aids, speech recognition

It is well documented that individuals who have hearing loss often complain of considerable difficulty understanding speech, especially in a background of noise. Recent advancements and improvements in digital hearing aid technology appear to have minimized this difficulty, as evidenced by the *subjective* reports provided by many self-assessment hearing aid outcome measures. However, much of the empirical research has not shown significant measurable advantages of digital technology over analog technology (Valente, Fabry, Potts, & Sandlin, 1998; Walden, Surr, Cord, Edwards, & Olson, 2000). This lack of evidence to support patients' subjective perceptions of digital technology suggests that their responses on self-assessment questionnaires may be artificially inflated. Therefore, a need exists to *objectively* document the improvement that hearing aid wearers actually experience. This is particularly true in the area of speech perception performance.

The earliest approaches to evaluating hearing aid benefit were comparison methods in which patient performance with two or three hearing aids was evaluated using traditional speech recognition tests, for example, Northwestern University Auditory Test No. 6 (NU-6) and Central Institute for the Deaf (CID) W-22 word lists (Carhart, 1946). However, such speech perception tests have long been criticized for not being sensitive enough to provide the information necessary to determine and define specific hearing aid benefit (Carhart, 1965; Mendel & Danhauer, 1997; Wiley, Stoppenbach, Feldhake, Moss, & Thordardottir, 1995). In the 1980s, with the advent of computerized probe microphone real ear technology, the challenge of developing scientifically based methods of selecting, evaluating, and fitting hearing aids became much easier (Northern, 1992). These objective probe microphone measurements were used to verify that the prescribed real ear gain of the hearing aid met desired targets. However, measuring insertion gain in this manner only provides information about the amount of real ear gain delivered by the hearing aid, and it does not necessarily supply any information about the patient's

1059-0889/07/1602-0118

speech-understanding ability in realistic listening situations. Unfortunately, many audiologists felt that matching hearing aid prescription targets using probe microphone technology was sufficient to verify hearing aid benefit, and therefore, such techniques have functionally served as a replacement for using speech recognition tests as measures of benefit. Consequently, many audiologists do not include speech recognition testing as part of the hearing aid evaluation process (Martin, Champlin, & Chambers, 1998).

Most recently, advancements in digital hearing aid technology have changed the hearing aid evaluation process once again. With extensive computer software now available to fit these digital instruments, not only are speech recognition test materials not being used but in some cases even real ear probe microphone verification techniques are not being utilized. Some manufacturers provide CD-ROMs with environmental sounds and unstandardized speech materials with their fitting software, but these stimuli, if used, are often played directly from a computer through uncalibrated speakers that are often not presented in a sound-treated room. Thus, the results from such evaluations are very difficult to quantify.

The terms *verification* and *validation* are often confused when used in the context of the hearing aid fitting process. Hearing aid verification techniques primarily focus on ways to confirm that the gain in the hearing aid matches the prescribed targets. Hearing aid verification is an important component of the hearing aid evaluation, but it does not evaluate whether the matched hearing aid targets are actually appropriate for the patient with regard to improvements in speech perception or whether the patient will benefit from such prescribed hearing aid gain. Therefore, recent research in hearing aid fitting has focused more on hearing aid validation techniques. Hearing aid validation refers to outcome measures designed to assess treatment efficacy (i.e., whether the hearing aids are beneficial). Weinstein (1997) has defined treatment efficacy as three different areas:

1. *Treatment effectiveness:* Do the hearing aids improve speech intelligibility in quiet and in noise or do they restore normal loudness perceptions?
2. *Treatment efficiency:* Are certain hearing aids or hearing aid settings/adjustments better than others for improving speech understanding?
3. *Treatment effects:* Does the use of hearing aids improve the patient's social or emotional well-being or his or her overall quality of life?

Because it is critically important for audiologists to demonstrate the outcomes of such treatments as hearing aids, much of our current clinical focus has shifted toward hearing aid validation. As a result, several self-assessment inventories have been developed in recent years in an effort to quantify patients' subjective perceptions of their hearing aid benefit (Cox & Alexander, 1995; Demorest & Erdman, 1986; Dillon, James, & Ginis, 1997; Newman & Weinstein, 1988; Turner, Humes, Bentler, & Cox, 1996; Walden, Demorest, & Hepler, 1984).

There are two different philosophies regarding how hearing aid validation techniques can document outcomes from the hearing aid fitting process: those that focus on subjective outcomes (i.e., using questionnaires and interviews to document the opinions and attitudes of the patient) and those that focus on objective outcomes (i.e., using empirical data to verify improvements in performance; Cox, 1999). Most studies of objectively measured hearing aid benefit have been conducted in a laboratory or clinical setting, which limits the generalization of those findings to more realistic listening environments (Cox, Alexander, & Gilmore, 1991). In recent years, several speech perception tests—for example, the Connected Speech Test (Cox, Alexander, & Gilmore, 1987), the Hearing in Noise Test (HINT; Nilsson, Soli, & Sullivan, 1994), the Speech-in-Noise test (Fikret-Pasa, 1993), and the Quick Speech-in-Noise test (QuickSIN; Etymotic Research, 2001; Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004)—have been developed with the goal of maximizing their face validity to provide a more accurate reflection of a listener's speech understanding. This project examined both objective and subjective outcome measures as a way to validate hearing aid benefit. The purpose was to determine whether some newly developed speech recognition materials were sensitive enough to demonstrate objective hearing aid benefit and whether such results would correlate well with patients' subjective perceptions of that benefit. Subjective outcomes seem to have become the "gold standard" to which hearing aid benefit results are compared. In this study, speech perception performance scores were compared with subjective self-assessment outcome measures to confirm improvements in performance.

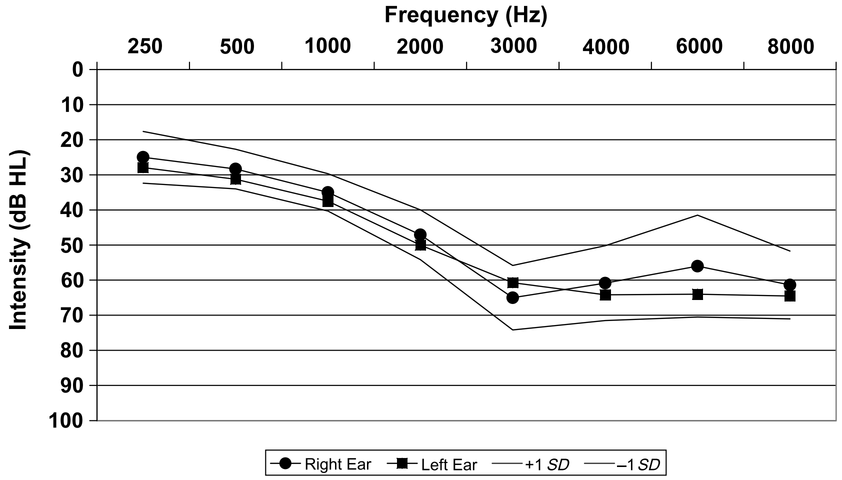
Method

Research Participants

Twenty-one adults ranging in age from 33 to 75 years (mean age = 65 years) participated in this study. All participants (9 men and 12 women) were hearing aid users (17 binaural, 4 monaural) who had used amplification from 6 months up to 6 years. All research participants had normal middle ear function at the time of testing as evidenced by Type A tympanograms bilaterally. They all had bilaterally symmetric sensorineural hearing losses of varying degree, with a mean pure-tone average of 35 dB in the right ear and 37 dB in the left ear as shown in Figure 1.

Table 1 displays detailed information about the research participants, including their age, gender, pure-tone averages, hearing aids and styles, and years of hearing aid use. The research participants were clients from the Memphis Speech and Hearing Center, so the hearing aid fitting procedures used were similar for all participants. All hearing aids fit were digital signal processing devices and were selected on the basis of the audiometric and lifestyle needs of each participant. Real ear probe microphone measurements were taken to verify the prescribed NAL-NL1 targets for each participant, and adjustments were made using an Audioscan Verifit and NOAH fitting software. Additional adjustments were made to the hearing aid fitting on the basis of feedback from the participants during the first month after fitting. Hearing aid orientation consisted of counseling regarding

Figure 1. Average audiometric thresholds for all research participants.



realistic expectations, gradual exposure from quiet to noisy situations, and care and use of the hearing aids. All participants were encouraged to take part in group adult audiologic rehabilitation sessions during the first 2–3 months of fitting. All participants signed an informed consent approved by The University of Memphis institutional review board for participation in this study, and basic ethical considerations were taken for the protection of the research participants throughout the project.

Stimuli

Three sentence tests of speech recognition designed to be used in a background of noise were administered to all participants in both unaided and aided conditions. The three tests were the Revised Speech Perception in Noise test (R-SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984), the QuickSIN (Etymotic Research, 2001; Killion et al., 2004), and the HINT (Nilsson et al., 1994). These tests were selected because they all have characteristics that enhance their face validity: (a) They all contain sentence stimuli, (b) they are all presented in a background of noise, and (c) they all have considerable standardization data available that suggest that they are valid tests of speech perception. The QuickSIN and the HINT were also selected because they are designed to measure the individual's signal-to-noise ratio (SNR) performance compared with normal performance. Thus, they measure *SNR loss,* which is defined as the decibel increase in SNR required by a person who is hearing impaired to understand speech in noise compared with someone with normal hearing (Etymotic Research, 2001).

The R-SPIN test was designed to be used in noise, and it takes into account the linguistic context of spoken utterances, which is known to be an important factor in normal spoken communication. It includes test words embedded in recognizable semantic contexts (e.g., "The dog chewed on a BONE"), allowing for enhanced predictability of the final word in the sentence. It also has items that are presented in semantically neutral contexts (e.g., "She wants to talk about the CREW") that are less predictable.

The QuickSIN consists of 12 lists of six sentences with five key words per sentence. The sentences are presented at prerecorded SNRs that decrease in 5-dB steps from 25 dB to 0 dB. Standardization data on the QuickSIN show that averaging the results from several lists improves the reliability of the score (i.e., it increases the number of test items). Thus, four six-sentence lists were administered per testing session (Etymotic Research, 2001).

Finally, the HINT, consisting of sentences presented in groups of 10 with competing noise, was administered. The HINT uses an adaptive method to measure the SNR at which the listener responds correctly 50% of the time. The adaptive method optimizes test efficiency by automatically adjusting the SNR for the upcoming trial on the basis of the response on the previous trial. The HINT battery consists of four tests with the speech signal presented at 0° in front of the listener: (a) speech in quiet (Quiet), (b) speech in noise with the noise source at 0° in front of the individual (Noise Front; NF), (c) speech in noise with the noise at 90° to the left (Noise Left; NL), and (d) speech in noise with the noise at 90° to the right (Noise Right; NR).

Instrumentation

All sentence testing was conducted in a sound-treated booth meeting American National Standards Institute's (1996) Standard S3.6-1996. All sentence stimuli were presented via a Grason-Stadler GSI-61 two-channel audiometer in the sound field with the participant facing the speaker at 0° azimuth approximately 1 m from the speaker. The R-SPIN stimuli were routed from a JVC (Model TD-W707) cassette tape deck and were presented at 50 dB HL, with the competing noise presented at a +6 dB SNR through the same speaker. Participants were instructed to repeat the last word in the sentence. The QuickSIN was routed from an

Table 1. Research participants.

Participant	Age (years)	Gender	Ear	PTA	HA make/model	HA style	Years of HA use
1	65	F	R	33	ReSound BT4	BTE	1
			L	37	ReSound BT4	BTE	
2	33	F	R	40	ReSoundBZ5	BTE	3
			L	45	ReSoundBZ5	BTE	
3	62	F	R	35	Phonak Claro 21 dAZ	ITE	2
			L	30	Phonak Claro 21 dAZ	ITE	
4	65	M	R	25	Oticon Adapto	BTE	1
			L	37	Oticon Adapto	BTE	
5	65	M	R	30	ReSound ED3	ITE	6
			L	27	ReSound ED3	ITE	
6	62	F	R	25	Phonak Claro 211 dAZ	BTE	1
			L	32	Phonak Claro 211 dAZ	BTE	
7	73	M	R	38	ReSound ED3	ITE	4
			L	55	ReSound ED3	ITE	
8	75	F	R	45	Oticon DigiFocus	BTE	1
			L	32	Oticon DigiFocus	BTE	
9	73	M	R	42	Phonak Claro 211	BTE	2
			L	40	Phonak Claro 211	BTE	
10	70	M	R	35	None	NA	2
			L	42	Oticon DigiFocus	ITE	
11	75	F	R	35	ReSound ED3	ITE	3
			L	43	ReSound ED3	ITE	
12	61	M	R	22	Widex Diva	ITE	0.5
			L	27	Widex Diva	ITE	
13	69	F	R	38	Siemens L5	ITC	0.5
			L	45	Siemens L5	ITC	
14	72	F	R	53	None	NA	1
			L	55	Oticon Atlas	ITE	
15	44	F	R	45	Oticon Adapto	ITE	1
			L	45	Oticon Adapto	ITE	
16	61	M	R	20	Phonak Perseo 23 dAZ	ITE	1.5
			L	22	Phonak Perseo 23 dAZ	ITE	
17	61	M	R	28	Widex Diva	ITE	0.5
			L	20	Widex Diva	ITE	
18	69	F	R	43	Phonak Perseo 23 dAZ	ITE	1
			L	35	None	NA	
19	72	F	R	50	None	NA	1
			L	53	Oticon Atlas	ITE	
20	44	F	R	40	Oticon Adapto	ITE	1
			L	45	Oticon Adapto	ITE	
21	61	M	R	20	Phonak Perseo 23 dAZ	ITE	1
			L	22	Phonak Perseo 23 dAZ	ITE	

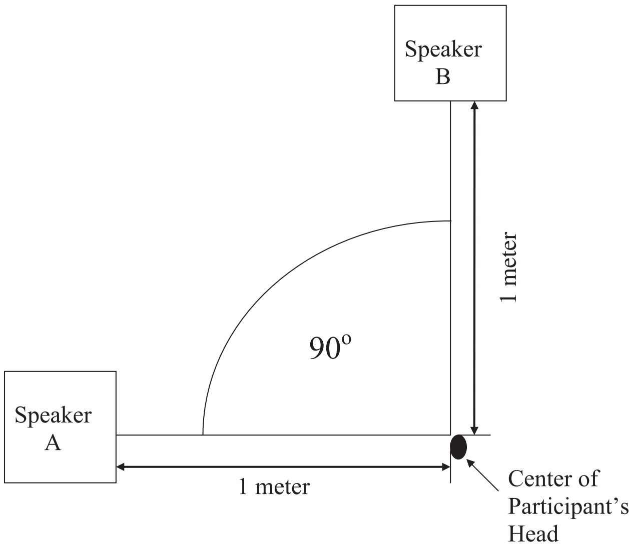
Note. PTA = pure-tone average; HA = hearing aid; F = female; M = male; R = right; L = left; BTE = behind the ear; ITE = in the ear; NA = not applicable; ITC = in the canal.

Onkyo (Model CDPC900) compact disc player and was presented at 70 dB HL (Etymotic Research, 2001), and the participants repeated the entire sentence. The HINT was administered using the procedures outlined in the *HINT for Windows Instrumentation Manual* produced by Maico Diagnostics (House Ear Institute, 2001). An adaptive up–down procedure was used to measure the reception threshold for speech (RTS), that is, the SNR at which the listener responded correctly 50% of the time. Equivalent lists for each of the sentence tests were randomly selected, and no lists were repeated for any participant throughout the duration of the study to minimize learning and practice effects. Each participant heard a different HINT list in each of the test conditions (Quiet, NF, NL, and NR) using the sound field set up displayed in Figure 2.

Subjective Questionnaire

In addition to the administration of the sentence recognition tests, the participants were asked to complete an outcome questionnaire assessing hearing aid benefit known as the Hearing Aid Performance Inventory (HAPI; Walden et al., 1984). This instrument was used because the HAPI specifically addresses situations relating to speech perception

Figure 2. Sound field speaker set-up for the Hearing in Noise Test (HINT) quiet and noise test conditions. The listener faced Speaker A for the Quiet (speech from Speaker A), Noise Front (speech and noise from Speaker A), and Noise Right (speech from Speaker A; noise from Speaker B) test conditions. The listener faced Speaker B for the Noise Left test condition (speech from Speaker B; noise from Speaker A). Adapted from the House Ear Institute's (2001) *HINT for Windows Instrumentation Manual—Version 6*.



performance in comparison with many of the other outcome measures currently available.

The participants responded to the 64 items on the HAPI by providing judgments about how much difficulty they had in a variety of listening situations both with and without their hearing aid(s). Participants responded to each listening situation using a 5-point scale. For the unaided condition, the responses were "Without my hearing aid, I …" (1) *always have difficulty,* (2) *generally have difficulty,* (3) *occasionally have difficulty,* (4) *seldom have difficulty,* and (5) *never have difficulty*. For the aided condition, the responses were "In this situation, my hearing aid …" (A) *is very helpful,* (B) *is helpful,* (C) *is very little help,* (D) *is no help,* and (E) *hinders performance*. For the purposes of data analysis in the aided condition, the responses given for A–E were converted to the numbers 5–1, respectively. The research participants were allowed as much time as they needed to complete the HAPI. Most took between 20 and 25 min to complete the questionnaire.

Scoring

For the R-SPIN test, only the last word in each sentence is repeated. Thus, the score for the R-SPIN is based on the number of correct words out of a total of 50 items per list. For the QuickSIN, the total number of correct key words from all six sentences in each list was used to calculate the SNR loss. The formula used was 25.5 – [total words correct] (Etymotic Research, 2001). The HINT measures the SNR at which the listener responds correctly 50% of the time (RTS). In quiet, the RTS was expressed as the level of the sentences in dB(A). When administered in noise, the RTS was expressed as SNR (the difference in dB between the sentence level and the noise level). A noise composite score was also computed from the front and side scores using the following formula: $[(2 \times \mathrm{NF} + \mathrm{NR} + \mathrm{NL})/4]$ (Nilsson et al., 1994). For the HAPI, the mean ratings for each item for the unaided and aided conditions were calculated.

Inter- and Intrajudge Scoring Reliability

To ensure accuracy in scoring the talk-back responses from the participants, interjudge and intrajudge scoring reliability was conducted on 30% of the data, using the following formula: [agreements/(agreements + disagreements)] × 100%. Inter- and intrajudge scoring reliability was 96% and 98%, respectively.

Results

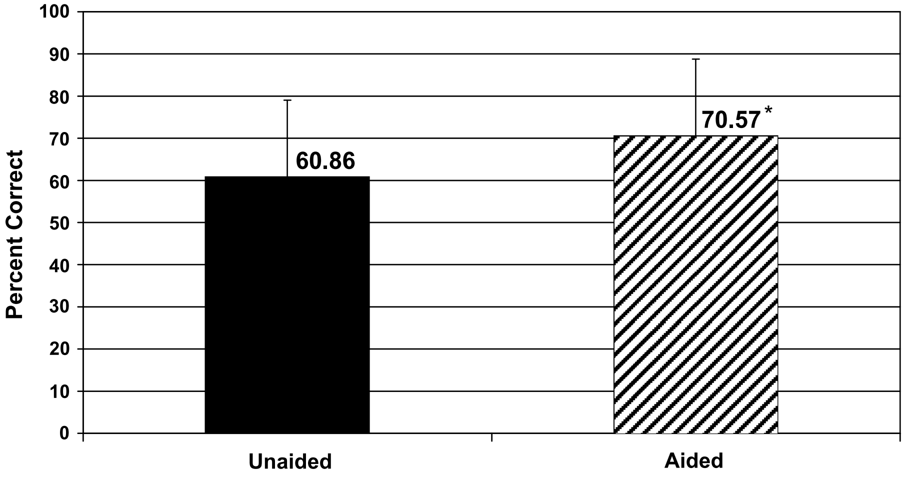
Objective Outcome Measures

R-SPIN. Overall performance on the R-SPIN was relatively poor in both the unaided and aided conditions. Figure 3 shows the mean percentage correct score for the R-SPIN in the unaided condition ($M = 60.86$) compared with the aided condition ($M = 70.57$). A one-way repeated measures analysis of variance (ANOVA) revealed that the mean R-SPIN score in the aided condition was significantly better than the mean R-SPIN score in the unaided condition, $F(1, 41) = 6.05$, $p = .023$. It should be noted that the scores on this test varied considerably as evidenced by the large standard deviations.

HINT. Figure 4 shows the HINT Quiet scores in the unaided and aided conditions. The HINT Quiet score is the dBA value in which the participant responded correctly 50% of the time. A one-way repeated measures ANOVA on the mean HINT Quiet scores revealed that aided performance was significantly better than unaided performance, $F(1, 41) = 39.69$, $p < .001$.

Figure 5 displays the mean HINT threshold scores and standard deviations in the three noise test conditions (NF, NR, and NL) as well as the composite score in the unaided and aided conditions. Some of the participants had considerable difficulty with some of the HINT noise conditions, making it difficult to calculate a score for each participant. Thus, because of the missing data points, paired t tests were performed on the HINT scores (reported as RTS) in the three noise conditions as well as for the noise composite. There were no significant differences measured between aided and unaided performance within any of the noise conditions or the noise composite condition. However, Figure 5 shows that there were statistically significant differences measured across all of the HINT noise conditions for both the unaided and aided scores. Unaided and aided HINT NF scores were significantly poorer (higher) than HINT NR scores: unaided, $t(16) = 4.65$, $p < .001$; aided, $t(16) = 7.03$, $p < .001$. Unaided and aided HINT NF scores were also significantly poorer (higher) than HINT NL scores: unaided, $t(18) = 14.67$, $p < .001$; aided,

Figure 3. Mean percentage of correct performance on the Revised Speech Perception in Noise test (R-SPIN) for unaided and aided conditions. Error bars represent +1 *SD*.



*[*p*=.023]

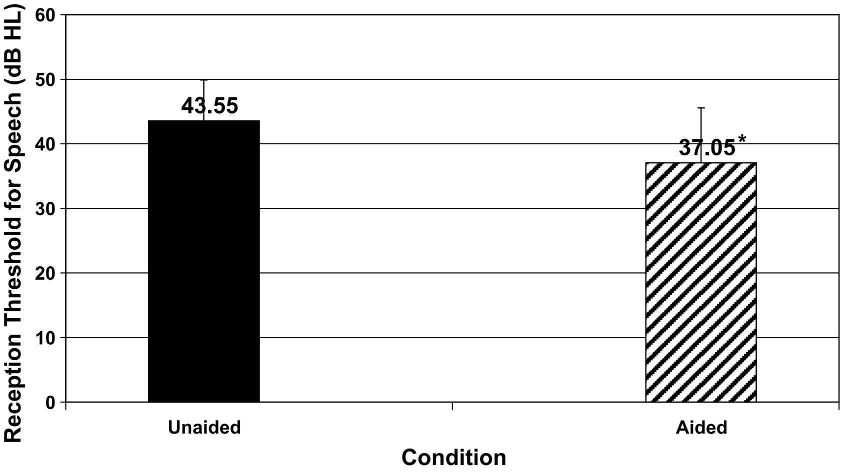
$t(15) = 13.14$, $p < .001$. HINT NR was significantly poorer than HINT NL for the unaided condition, $t(16) = 3.00$, $p = .008$, and the aided condition, $t(17) = 2.49$, $p = .025$.

QuickSIN. For the QuickSIN, Figure 6 shows that scores generally improved as the SNR improved; all participants had the most difficulty in the 0-dB SNR condition for both unaided and aided performance. Paired *t* tests revealed that aided QuickSIN scores were significantly better than unaided scores for all SNRs except 0-dB SNR ($p < .05$). See Table 2 for *t* values. SNR loss was calculated by defining the decibel increase in SNR required by a person who is hearing impaired to understand speech in noise compared with someone with normal hearing. Figure 7 shows that the mean SNR loss for the aided condition ($M = 5.75$) was significantly better than in the unaided condition ($M = 7.90$), $t(17) = 2.40$, $p = .03$.

Subjective Outcome Measures

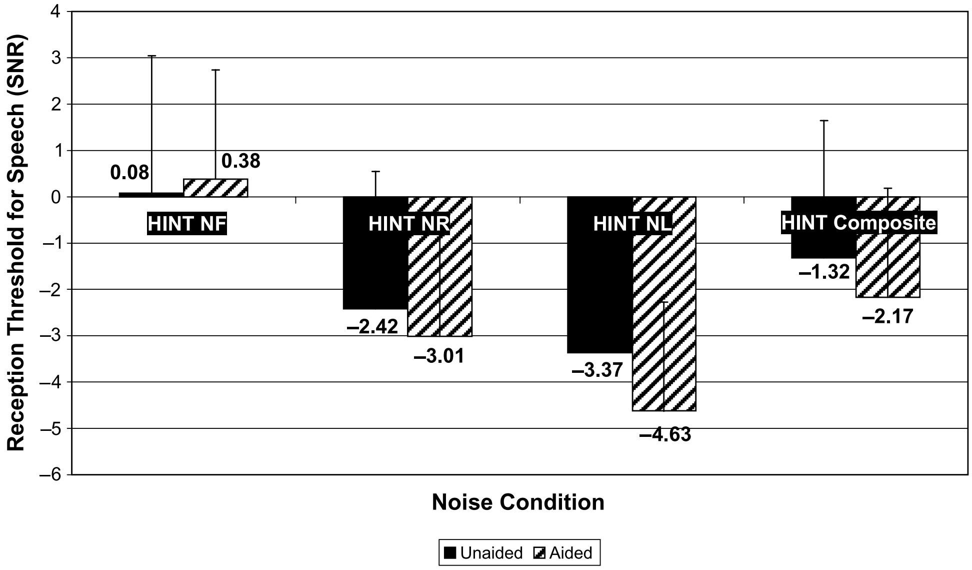
The 64 items of the HAPI were divided into five categories based on the basic content of each question. These categories included the following: perception of environmental sounds in quiet and noise (ENV Q&N), conversation in quiet situations with familiar talkers (Quiet F),

Figure 4. Mean HINT Quiet threshold (in dBA; reception threshold for speech) for the unaided and aided conditions. Error bars represent +1 *SD*.



*[*p*<.001]

Figure 5. Mean HINT reception threshold for speech (in dB) signal-to-noise ratio (SNR) for the three test noise conditions—Noise Front (HINT NF), Noise Right (HINT NR), Noise Left (HINT NL)—and the noise composite for the unaided and aided conditions. Error bars represent +1 *SD*.



conversation in quiet situations with unfamiliar talkers (Quiet UF), conversation in noisy situations with familiar talkers (Noise F), and conversation in noisy situations with unfamiliar talkers (Noise UF). Figure 8 displays the mean ratings for these categories on the HAPI for the unaided and aided conditions. The more positively the participant responded to a specific item, the higher the number of the rating. Separate one-way repeated measures ANOVAs for each questionnaire category (ENV Q&N, Quiet F, Quiet UF, Noise F, and Noise UF) revealed that all aided responses were significantly better than the unaided responses in all categories except for ENV Q&N. See Table 3 for *F* values.

Correlations Among Objective and Subjective Outcome Measures

Pearson product–moment correlations were conducted to determine whether there were any significant relationships among the subjective measures (five categories of HAPI responses) and objective measures (R-SPIN; HINT [Quiet, NF, NR, NL]; QuickSIN SNR Loss). Tables 4, 5, and 6 display the significant correlations measured among the variables. Table 4 shows the significant correlations among the subjective outcome measures indicating that ENV Q&N and Quiet F were highly correlated with all of the other

Figure 6. Mean number of correct responses on the Quick Speech-in-Noise test (QuickSIN) at each SNR (25, 20, 15, 10, 5, and 0 dB) for both the aided and unaided conditions. Error bars represent +1 *SD*.

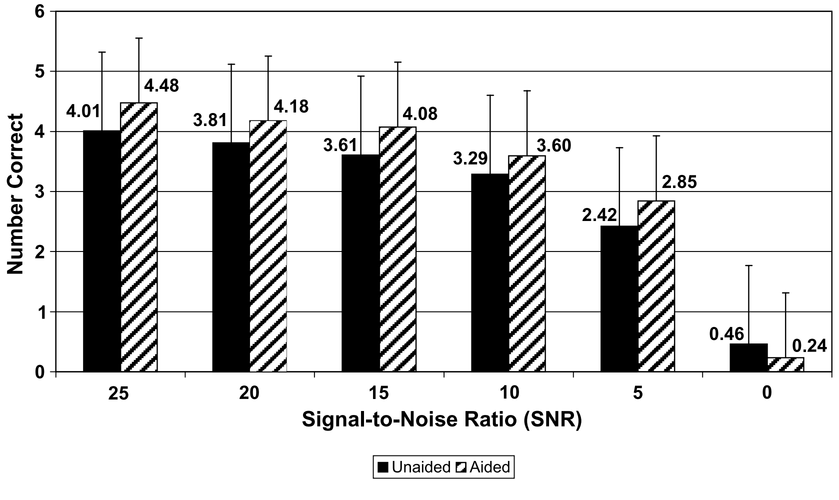


Table 2. Paired *t* tests for unaided and aided QuickSIN SNR scores.

SNR	*t*	*df*	*p*
25 dB	–2.23	17	.04[a]
20 dB	–2.22	17	.04[a]
15 dB	–2.20	17	.04[a]
10 dB	–2.17	17	.04[a]
5 dB	–2.15	17	.05[a]
0 dB	1.35	17	.20
SNR Loss	2.40	17	.03[a]

Note. QuickSIN = Quick Speech-in-Noise test; SNR = signal-to-noise ratio.

[a] Denotes statistically significant difference.

subjective variables in the unaided condition and with fewer HAPI categories in the aided condition. Quiet UF was highly correlated with Quiet F in the unaided condition (.65, $p < .001$) and in the aided condition (.83, $p < .0001$). Noise F was correlated with Quiet F and Quiet UF. These correlations were higher in the unaided condition (.89, $p < .0001$ UF; .71, $p < .0003$ F) compared with the aided condition (.54, $p < .01$ UF; .47, $p < .03$ F). Noise UF was also correlated with Quiet F (.67, $p < .0009$) and Quiet UF (.91, $p < .0001$) in the unaided condition and with Quiet UF (.51, $p < .01$) and Noise F (.91, $p < .0001$) in the aided condition.

Table 5 displays all of the statistically significant correlations measured among the objective measures. Examination of this table indicates that for the unaided condition, R-SPIN was significantly correlated with all of the HINT tests and with QuickSIN SNR Loss. In addition, all of the HINT conditions were significantly correlated with QuickSIN SNR Loss except HINT NF. For the aided condition, there were slightly fewer significant correlations among the objective variables than in the unaided condition.

Table 6 combines the information provided in Tables 4 and 5 and displays the significant correlations measured between the subjective and objective outcome measures. In the unaided condition, Quiet F was significantly correlated with all of the objective measures. In addition, Quiet UF was negatively correlated with HINT Quiet (–.48, $p < .02$) and HINT NR (–.53, $p < .02$), indicating that as research participants rated their abilities in quiet conversation high, their HINT scores decreased (improved). In the aided condition, Noise F and Noise UF were significantly correlated with all of the objective variables except HINT NL. Also, Quiet UF was significantly correlated with R-SPIN (.55, $p < .01$), suggesting that as subjective ratings increased, so did the R-SPIN scores. The Quiet UF category was also negatively correlated with QuickSIN SNR Loss (–.66, $p < .001$), indicating that as subjective ratings improved, the SNR loss decreased (improved).

A multiple regression analysis was conducted on the data to determine whether any of the objective or subjective measures could predict performance on any other measure. A few of the variables were found to be significant. In the aided condition, the dependent variable Noise UF was predicted from the linear combination of the independent variables of R-SPIN and HINT Quiet ($r^2 = .74$, $p = .02$). In the unaided condition, the dependent variable Quiet F was predicted from the linear combination of the independent variables of HINT Quiet and HINT NF ($r^2 = .94$, $p = .003$). In addition, Quiet UF was predicted from R-SPIN ($r^2 = .92$, $p = .006$), and ENV Q&N was predicted from R-SPIN and HINT NR ($r^2 = .62$, $p = .04$).

Discussion

This investigation studied both objective and subjective outcome measures as a way to validate hearing aid benefit. Participants' unaided and aided performance on sentence recognition tests in noise was measured to demonstrate objective hearing aid benefit. These objective results were then compared with the participants' subjective perceptions of that benefit using the HAPI questionnaire.

Figure 7. Mean QuickSIN SNR loss for the unaided and aided conditions. Error bars represent +1 *SD*.

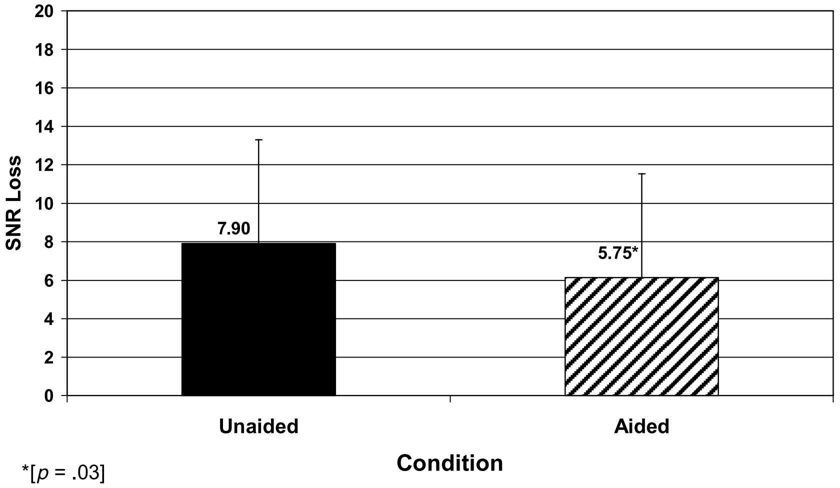
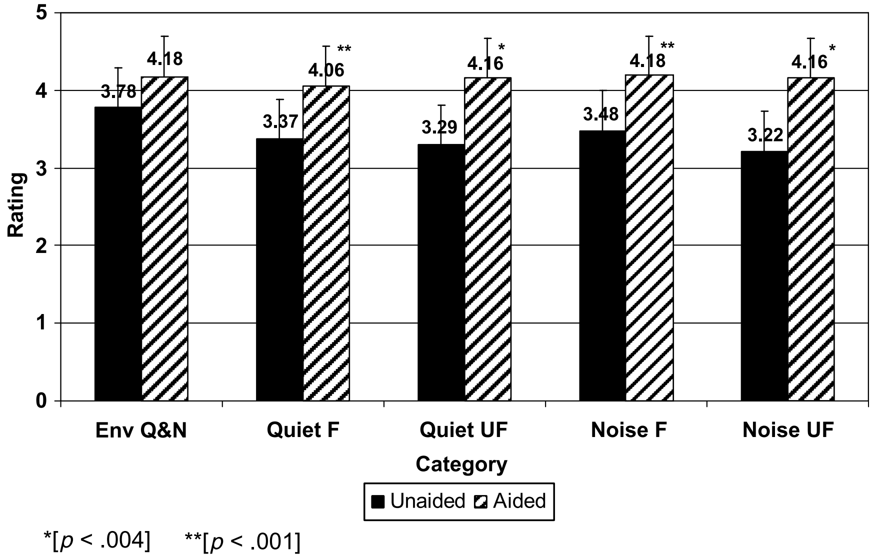


Figure 8. Mean subjective ratings on the Hearing Aid Performance Inventory (HAPI) as a function of listening category for the unaided and aided conditions. Env Q&N = perception of environmental sounds in quiet and noise; Quiet F = conversation in quiet situations with familiar talkers; Quiet UF = conversation in quiet situations with unfamiliar talkers; Noise F = conversation in noisy situations with familiar talkers; Noise UF = conversation in noisy situations with unfamiliar talkers. Error bars represent +1 *SD*.



Objective Outcome Measures

Objective assessment of speech perception performance using the R-SPIN, HINT, and QuickSIN produced varied results in both unaided and aided conditions. Significant improvements in performance between unaided and aided conditions were measured for the R-SPIN, HINT Quiet, and QuickSIN SNR Loss. Despite the relatively poor performance on the R-SPIN in both conditions, this objective measure was sensitive enough to reveal a statistically significant difference between the unaided and aided conditions. The HINT Quiet threshold also showed a statistically significant improvement in the aided compared with the unaided condition. Given that this is a threshold measure, documented improvement in threshold from unaided to aided is to be expected. Finally, the SNR loss calculated by the QuickSIN revealed significantly better aided SNR loss compared with the unaided SNR loss, even though comparisons within each SNR of the QuickSIN revealed no significant differences between unaided and aided conditions.

Unfortunately, objective speech perception performance in the HINT noise conditions (NF, NR, and NL) was not as sensitive as evidenced by the lack of statistically significant differences measured between unaided and aided performance. This lack of sensitivity could be due to the fact that some of the participants had considerable difficulty with the stimuli presented in the HINT noise conditions, thus

Table 3. One-way repeated measures analysis of variance for unaided and aided HAPI scores.

Category	*F*	*df*	*p*
ENV Q&N	2.73	20	.114
QUIET UF	10.52	20	.004[a]
QUIET F	28.34	20	<.001[a]
NOISE UF	13.29	20	.002[a]
NOISE F	16.28	20	<.001[a]

Note. HAPI = Hearing Aid Performance Inventory; ENV Q&N = perception of environmental sounds in quiet and noise; Quiet UF = conversation in quiet situations with unfamiliar talkers; Quiet F = conversation in quiet situations with familiar talkers; Noise UF = conversation in noisy situations with unfamiliar talkers; Noise F = conversation in noisy situations with familiar talkers.

[a]Denotes statistically significant difference.

Table 4. Pearson product–moment correlations among the subjective outcome measures.

HAPI category	Unaided	Aided
ENV Q&N	Quiet UF (.82; $p < .0001$) Quiet F (.79; $p < .0001$) Noise UF (.87; $p < .0001$) Noise F (.83; $p < .0001$)	Quiet UF (.42; $p < .05$) Quiet F (.47; $p < .03$)
Quiet UF	Quiet F (.65; $p < .001$)	Quiet F (.83; $p < .0001$)
Quiet F	ENV Q&N (.79; $p < .0001$) Quiet UF (.65; $p < .001$) Noise UF (.67; $p < .0009$) Noise F (.71; $p < .0003$)	ENV Q&N (.47; $p < .03$) Quiet UF (.83; $p < .0001$) Noise F (.47; $p < .03$)
Noise UF	Quiet UF (.91; $p < .0001$) Quiet F (.67; $p < .0009$)	Quiet UF (.51; $p < .01$) Noise F (.91; $p < .0001$)
Noise F	Quiet UF (.89; $p < .0001$) Quiet F (.71; $p < .0003$)	Quiet UF (.54; $p < .01$) Quiet F (.47; $p < .03$)

Note. Values in parentheses indicate correlation coefficients followed by *p* values.

Table 5. Pearson product–moment correlations among the objective outcome measures.

Test	Unaided	Aided
R-SPIN	HINT Quiet (–.68; *p* < .0006)	HINT Quiet (–.81; *p* < .0001)
	HINT NF (–.66; *p* < .002)	HINT NF (–.68; *p* < .002)
	HINT NR (–.78; *p* < .0002)	HINT NR (–.59; *p* < .005)
	HINT NL (–.69; *p* < .0005)	HINT NL (–.46; *p* < .05)
	SNR Loss (–.90; *p* < .0001)	SNR Loss (–.92; *p* < .0001)
HINT Quiet	HINT NR (.82; *p* < .0001)	HINT NR (.51; *p* < .02)
	HINT NL (.61; *p* < .003)	
	SNR Loss (.78; *p* < .0001)	SNR Loss (.52; *p* < .01)
HINT NF	HINT NR (.77; *p* < .0002)	HINT NR (.59; *p* < .01)
	HINT NL (.87; *p* < .0001)	HINT NL (.70; *p* < .002)
		SNR Loss (.64; *p* < .005)
HINT NR	HINT NL (.86; *p* < .0001)	HINT NL (.77; *p* < .0001)
	SNR Loss (.71; *p* < .004)	SNR Loss (.52; *p* < .01)
HINT NL	SNR Loss (.52; *p* < .02)	
QuickSIN SNR Loss	R-SPIN (–.90; *p* < .0001)	R-SPIN (–.92; *p* < .0001)
	HINT Quiet (.78; *p* < .0001)	HINT Quiet (.52; *p* < .01)
		HINT NF (.64; *p* < .005)
	HINT NR (.71; *p* < .004)	HINT NR (.52; *p* < .01)
	HINT NL (.52; *p* < .02)	

Note. Values in parentheses indicate correlation coefficients followed by *p* values. R-SPIN = Revised Speech Perception in Noise test; HINT = Hearing in Noise Test; NF = Noise Front; NR = Noise Right; NL = Noise Left.

making this test too difficult to yield useful results for them. In addition, the relatively small number of participants in this sample reduced the power of the test; a larger subject population might yield more power and make the data more useful.

Nonetheless, the results from the objective outcome measures by themselves suggest that some sentence recognition tests are better than others in assessing objective improvements in speech perception performance with hearing aids. Specifically, the R-SPIN, the HINT Quiet threshold,

Table 6. Pearson product–moment correlations among both subjective and objective outcome measures.

HAPI category	Unaided	Aided
ENV Q&N	R-SPIN (.58; *p* < .005)	
	HINT Quiet (–.59; *p* < .004)	
		HINT NR (.45; *p* < .04)
		HINT NL (.54; *p* < .02)
Quiet UF	HINT Quiet (–.48; *p* < .02)	
	HINT NR (–.53; *p* < .02)	
		R-SPIN (.55; *p* < .01)
		SNR Loss (–.66; *p* < .001)
Quiet F	R-SPIN (.66; *p* < .001)	
	HINT Quiet (–.59; *p* < .004)	
	HINT NR (–.62; *p* < .007)	
	HINT NL (–.47; *p* < .03)	
	SNR Loss (–.50; *p* < .02)	
Noise UF		R-SPIN (.62; *p* < .002)
		HINT Quiet (–.80; *p* < .0001)
		HINT NF (–.47; *p* < .05)
		HINT NR (–.50; *p* < .02)
		SNR Loss (–.75; *p* < .0001)
Noise F		R-SPIN (.62; *p* < .002)
		HINT Quiet (–.67; *p* < .0008)
		HINT NF (–.54; *p* < .02)
		HINT NR (–.49; *p* < .02)
		SNR Loss (–.73; *p* < .0001)

Note. Values in parentheses indicate correlation coefficients followed by *p* values.

and the QuickSIN SNR Loss appear to be the most sensitive of the tests used here to objectively assess speech perception performance in noise in both unaided and aided conditions. These objective measures are the most robust and can provide very useful information regarding a person's speech perception in noise comparing unaided and aided performance.

Subjective Outcome Measures

Analysis of the responses on the HAPI revealed that the aided responses averaged over four of the five categories (Quiet F, Quiet UF, Noise F, Noise UF) were significantly better than the unaided responses, suggesting that the HAPI was able to document improvements in aided versus unaided performance. Thus, this outcome measure was successful at assessing overall subjective hearing aid benefit. It is possible that no significant differences were measured for the ENV Q&N category because the research participants in this study were only moderately impaired and may have had little trouble perceiving environmental stimuli without their hearing aids anyway.

The correlations measured among the various categories of the HAPI (see Table 4) provided more revealing information regarding the usefulness of this questionnaire as an outcome measure. The items in the ENV Q&N and Quiet F categories were significantly correlated with all of the other HAPI categories in the unaided condition, suggesting that this instrument was sensitive to the subjective responses provided by the research participants. There were fewer significant correlations among the aided categories, suggesting that the specific items–categories within the questionnaire may have had an influence on some of the results.

The large number of significant correlations measured among the objective tests (see Table 5) reinforces the fact that the R-SPIN, HINT Quiet, and QuickSIN SNR Loss are sensitive objective measures of speech perception performance in noise. More data may need to be collected on the HINT noise conditions to make a final determination of the usefulness of these subtests in hearing aid evaluations.

Table 6 reveals the most interesting findings of all the correlations, as it displays the significant correlations measured between the objective and subjective test measures. Recall that significant correlations were measured between HAPI ratings Quiet F and all of the objective measures except HINT NF in the unaided condition. In addition, HAPI ratings for Noise F and Noise UF were significantly correlated with all of the objective measures except HINT NL. This large number of statistically significant correlations among the HAPI ratings and the objective measures indicates that as HAPI ratings improved, speech perception scores improved as well. These findings are most significant because they verify that both subjective and objective outcome measures for these cases were equally able to document improvements in speech perception performance among these hearing aid wearers.

Even though this study was not able to show significant correlations among all of the subjective and objective measures, the findings reported here are still encouraging and warrant further study. The fact that HAPI ratings for conversations in quiet and in noise were highly correlated with the R-SPIN, the HINT Quiet thresholds, and the QuickSIN SNR Loss suggests an encouraging trend documenting not only the relationship between these subjective and objective measures but also the sensitivity of these instruments in validating performance.

Perhaps the strongest statement of the relationship between the objective and subjective outcome measures would be through significant regression analyses, because regression analyses could show how one measure might predict performance in another measure. The results from the multiple regression analysis that was conducted on these data are encouraging and warrant further study. Additional data would increase the power of the correlations and the regressions and might yield further significant findings between the objective and subjective outcome measures.

Subjective outcomes seem to have become the "gold standard" to which hearing aid benefit results are compared. The results of this study better define the relationship between the scores obtained on objective sentence tests and the subjective responses on the HAPI. Objective documentation of subjective impressions is essential for determining the efficacy of treatment outcomes in hearing aid fitting. Such findings can have the potential for strong clinical impact. That is, objective documentation that some speech recognition test materials are sensitive measures of speech perception performance should support the use of such test materials in the hearing aid evaluation process.

The results of this study suggest that collecting objective outcome data from the R-SPIN, the HINT Quiet threshold, and the QuickSIN SNR Loss, along with subjective outcome data from the HAPI, can help to quantify the most important and desirable benefit of hearing aids—that of improved speech perception performance. If such improvement is not documented both subjectively and objectively in the hearing aid evaluation process, then it is difficult to verify that improvement has occurred. The results of this study have begun to define the relationship between objective and subjective outcome measures in an attempt to better define true hearing aid benefit.

Conclusion

Objective and subjective outcome measures were used in this study to validate hearing aid benefit. Documented improvements in aided versus unaided performance were found for the R-SPIN, the HINT Quiet threshold, and the QuickSIN SNR Loss for the objective outcome measures. For the subjective outcome measures, the HAPI was able to document hearing aid benefit for all of the subjective categories of the questionnaire. However, correlations among the objective and subjective outcomes were not significant in all cases, suggesting that the sensitivity of the HAPI and the sentence recognition tests used here vary according to specific items and procedures. In addition, the small sample size of this study limits any broad conclusions that can be made from these results.

The most noteworthy correlations between the subjective and objective outcome measures documented that HAPI ratings improved as the R-SPIN, the HINT Quiet threshold, and the QuickSIN SNR Loss improved. This finding is quite important because it validates both subjective and objective improvements in speech perception performance. Ultimately, objective measures taken together with subjective measures can provide the strongest measures of sensitivity.

The findings presented here remind us that the hearing aid evaluation process could benefit from including both objective and subjective outcome measures. Specifically, the R-SPIN, the HINT Quiet threshold, and the QuickSIN SNR Loss could be included as objective speech recognition tests as an integral part of the hearing aid evaluation process. These objective tests along with data obtained from subjective questionnaires, such as the HAPI, will strengthen the hearing aid test battery. If subjective outcomes are to become the "gold standard" to which hearing aid benefit is compared, it is clear from the results of this study that the addition of some objective outcome measures can help validate the patients' subjective impressions and better determine the efficacy of treatment outcomes in hearing aid fitting. Objective and subjective quantification of improved speech perception performance must be documented in the hearing aid evaluation process to verify that improvement has occurred. Ultimately, such results may help to provide more standardization to the hearing aid evaluation process across clinics.

Acknowledgments

Portions of this article were presented at the 27th International Congress of Audiology, Phoenix, Arizona, September 2004, and the American Speech-Language-Hearing Association Convention, Philadelphia, Pennsylvania, November 2004. This work was funded through the 2001 Audiology Outcomes Grant from the American Speech-Language-Hearing Foundation. Appreciation is expressed to Ginger A. Gray and Bridget G. Lee for their assistance in developing the protocols and collecting data. Thanks also to the clinical faculty and students from the Memphis Speech and Hearing Center for their assistance in the recruitment of research participants.

References

American National Standards Institute. (1996). *Specification for audiometers* (ANSI S3.6-1996). New York: Author.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research, 27,* 32–48.

Carhart, R. (1946). Speech reception in relation to pure tone loss. *Journal of Speech and Hearing Research, 11,* 97–108.

Carhart, R. (1965). Problems in the measurement of speech discrimination. *Archives of Otolaryngology, 82,* 253–260.

Cox, R. (1999). Measuring hearing aid outcomes: Part 1. *Journal of the American Academy of Audiology, 10,* Editorial.

Cox, R., & Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing, 16,* 176–183.

Cox, R., Alexander, G. C., & Gilmore, C. (1987). Development of the Connected Speech Test (CST). *Ear and Hearing, 8,* 119S–126S.

Cox, R., Alexander, G. C., & Gilmore, C. (1991). Objective and self-report measures of hearing aid benefit. In G. A. Studebaker, F. H. Bess, & L. B. Beck (Eds.), *The Vanderbilt Hearing Aid Report II* (pp. 201–213). Parkton, MD: York Press.

Demorest, M., & Erdman, S. (1986). Scale composition and item analysis for the communication profile of hearing aid benefit. *Journal of Speech and Hearing Research, 29,* 515–535.

Dillon, H., James, A., & Ginis, J. (1997). Client Oriented Scale of Improvement (COSI) and its relationship to several other measures of benefit and satisfaction provided by hearing aids. *Journal of the American Academy of Audiology, 8,* 27–43.

Etymotic Research. (2001). *QuickSIN: Speech-in-Noise Test* (Version 1.3). Elk Grove Village, IL: Author.

Fikret-Pasa, S. (1993). *The effects of compression ratio on speech intelligibility and quality.* Unpublished doctoral dissertation, Northwestern University, Chicago.

House Ear Institute. (2001). *HINT for Windows instrumentation manual—Version 6*. Eden Prairie, MN: Maico Diagnostics.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 116,* 2395–2405.

Martin, F. N., Champlin, C. A., & Chambers, J. A. (1998). Seventh survey of audiometric practices in the United States. *Journal of the American Academy of Audiology, 9,* 95–104.

Mendel, L. L., & Danhauer, J. L. (1997). Test administration and interpretation. In L. L. Mendel & J. L. Danhauer (Eds.), *Audiologic evaluation and management and speech perception assessment* (1st ed., pp. 15–57). San Diego, CA: Singular.

Newman, C., & Weinstein, B. E. (1988). The Hearing Handicap Inventory for the Elderly as a measure of hearing aid benefit. *Ear and Hearing, 9,* 81–85.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America, 95,* 1085–1099.

Northern, J. L. (1992). Introduction to computerized probe microphone real ear measurements in hearing aid evaluation procedures. In H. G. Mueller, D. B. Hawkins, & J. L. Northern (Eds.), *Probe microphone measurements: Hearing aid selection and assessment* (1st ed., pp. 1–19). San Diego, CA: Singular.

Turner, C., Humes, L., Bentler, R., & Cox, R. (1996). A review of past research on changes in hearing aid benefit over time. *Ear and Hearing, 17,* 14S–29S.

Valente, M., Fabry, D. A., Potts, L. G., & Sandlin, R. E. (1998). Comparing the performance of the Widex SENSO digital hearing aid with analog hearing aids. *Journal of the American Academy of Audiology, 9,* 342–360.

Walden, B. E., Demorest, M., & Hepler, E. (1984). Self report approach to assessing benefit derived from amplification. *Journal of Speech and Hearing Research, 27,* 49–56.

Walden, B. E., Surr, R. K., Cord, M. T., Edwards, B., & Olson, L. (2000). Comparison of benefits provided by different hearing aid technologies. *Journal of the American Academy of Audiology, 11,* 540–560.

Weinstein, B. E. (1997). Introduction: Customer satisfaction with and benefit from hearing aids. *Seminars in Hearing, 18,* 3–5.

Wiley, T. L., Stoppenbach, D. T., Feldhake, L. J., Moss, K. A., & Thordardottir, E. T. (1995). Audiologic practices: What is popular versus what is supported by evidence. *American Journal of Audiology, 4,* 26–34.

Received October 13, 2006
Revision received March 8, 2007
Accepted May 9, 2007
DOI: 10.1044/1059-0889(2007/016)

Contact author: Lisa Lucks Mendel, School of Audiology and Speech-Language Pathology, The University of Memphis, 807 Jefferson Avenue, Memphis, TN 38105. E-mail: lmendel@memphis.edu.