J Am Acad Audiol 3: 242–254 (1992)

Predictability and Reliability of Hearing Aid Benefit Measured Using the PHAB

Robyn M. Cox*
Izel M. Rivera*†

Abstract

This investigation explored the extent to which self-assessed hearing aid benefit measured by the Profile of Hearing Aid Benefit (PHAB) could be predicted from adaptation to hearing loss and/or from communication difficulties reported without amplification. Adaptation to hearing loss was measured using 13 scales of the Communication Profile for the Hearing Impaired (CPHI). These were combined to produce three composite CPHI scores. Results from 58 experienced hearing aid wearers indicated that benefit was significantly related to magnitude of unaided difficulties for all seven PHAB subscales. In addition, one of the three composite CPHI scores contributed to benefit prediction for the two sound perception subscales of the PHAB. Test-retest reliability of PHAB subscale scores was evaluated for 28 subjects using correlations and difference distributions. Reliability was found to be consistent with previous studies but modest. Critical differences were large compared with the anticipated size of benefit differences due to, for example, different hearing aid prescriptions. It is concluded that the PHAB is best suited for group research: when used for individual subjects, the PHAB may not be sensitive enough to detect important differences between hearing aid conditions.

Key Words: Hearing aids, self-assessment, prediction of benefit, reliability

Because the judgment of the wearer is the ultimate criterion of success in a hearing aid fitting, and because no laboratory or clinical procedure has emerged that accurately predicts this judgment, the use of subjective assessment to measure hearing aid benefit has much intuitive appeal. Subjectively assessed hearing aid benefit can play several roles in rehabilitation. It may be used to secure a global (single-figure) measure of the outcome of a hearing aid fitting to determine whether sufficient overall benefit has been achieved. Alternatively, subjective benefit may be used analytically to quantify specific problem areas in an existing hearing aid fitting: identified problem areas can then be addressed with adjustments to amplification or appropriate counseling. The literature reveals that subjective assessments of benefit also figure importantly in providing empirical validation or comparison of laboratory-derived strategies for hearing aid fittings. Finally, subjective assessment procedures can produce criterion data that are used to validate less cumbersome clinical or laboratory approaches to benefit measurement.

Despite this impressive array of potential applications, clinicians and researchers alike may be slow to embrace subjective benefit data. It seems possible, for example, that such data may be unreliable as a result of day-to-day fluctuations in mood, health, or recent experiences. Moreover, opinions may be biased by the individual's perceptions of, and personal adaptation to, the hearing loss or the hearing aid. Responses also may be affected by patients' desire to please the audiologist by providing a positive report. These types of concerns have been intensified by recent reports of clearly invalid or unreliable subjective benefit assessments obtained with particular procedures (Green et al, 1989; McClymont et al, 1991). This article describes work undertaken to evaluate some of these issues with respect to the Profile of Hearing Aid Benefit (PHAB). This self-assessment tool was designed to measure the

---

*Department of Veterans Affairs Medical Center and †Memphis State University, Memphis, TN
Reprint requests: Robyn M. Cox, Memphis Speech and Hearing Center, 807 Jefferson Ave., Memphis, TN 38105

proportion of time that a hearing aid: (1) improves speech communication in situations that are frequently encountered in daily life and, (2) increases the aversiveness or decreases the quality of sounds.

The PHAB is a 66-item inventory that uses the same items as the Profile of Hearing Aid Performance (Cox and Gilmore, 1990). Each item is a statement, such as "I have to ask people to repeat themselves in one-on-one conversation in a quiet room." The respondent's task is to indicate the proportion of time that the statement is true, using a seven-point scale as follows: always (99%), almost always (87%), generally (75%), half-the-time (50%), occasionally (25%), seldom (12%), and never (1%). Each response choice includes both a descriptor and a percentage. Responses are scored in terms of the percentage. Each item is answered twice, once for "without my hearing aid" and again for "with my hearing aid." Hearing aid benefit is defined as the difference between the two responses.

The PHAB was designed to be scored in terms of seven subscales or four scales. However, because the scale scores have not been found to be very useful, only the subscales were used in the present study. The test items and details of subscale development may be found in Cox and Gilmore (1990). Briefly, the subscales are:

Familiar Talkers (FT). Seven items describing communication under relatively easy listening conditions with persons whose voices are known.

Ease of Communication (EC). Seven items describing the effort involved in communication under relatively easy listening conditions.

Reverberation (RV). Nine items describing speech understanding in moderately reverberant rooms.

Reduced Cues (RC). Nine items describing communication without visual cues or when intensity is low.

Background Noise (BN). Sixteen items describing speech understanding in the presence of multitalker babble or other environmental competing noise.

Aversiveness of Sounds (AV). Twelve items describing negative reactions to environmental sounds.

Distortion of Sounds (DS). Six items describing the quality of voices and other sounds.

Cox, Gilmore, and Alexander (1991) reported an initial evaluation of the psychometric properties of the subscales of the PHAB, including means, standard deviations, and internal consistency reliability coefficients. The present work was designed, in part, to validate these statistics. Also, we evaluated the reliability of the subscales and computed critical differences for assessing the significance of changes in individual scores.

A primary question in this work was the extent to which an individual's self-assessed benefit could be predicted, based on the magnitude of unaided difficulties he or she reports and/or the individual's adaptation to hearing loss. It is obvious that persons who report a small proportion of communication problems without a hearing aid will realize little subjective benefit from amplification. However, it is less clear whether individuals who report a large proportion of communication problems without a hearing aid usually net relatively large subjective benefit when a hearing aid is acquired.

In addition, it seems reasonable to postulate a relationship between subjective benefit and adaptation to the hearing loss. For example, a hearing-impaired individual who has adapted poorly to communication problems (denies them, blames others, or is often angry or depressed) might tend to belittle the assistance provided by a hearing aid. Similarly, one might argue that a person who has developed habits that facilitate successful communication (such as asking for repetitions, optimizing seating, etc.) would be more likely to detect significant benefits attendant upon hearing aid use. If this type of relationship does exist, valid interpretation of subjective benefit data would require simultaneous consideration of personal adaptation data.

A literature search revealed only one reported investigation that reflects upon the relationship between subjectively estimated hearing aid benefit and the attitudes of the hearing aid wearer. Brooks (1989) measured benefit in terms of reported hours of daily use of the hearing aid. Using this metric, he determined that benefit was greater for individuals who believed that, because of their hearing loss, they were hard to talk to, missing important sounds, avoiding new people, and eliciting impatient reactions from friends and family. Furthermore, benefit was less for individuals who denied that they had a significant hearing problem. This study generally supported a hypothesis that adaptation to hearing loss may be related to subjective hearing aid benefit.

EXPERIMENT 1

Predictability of Benefit

This study examined the extent to which subjective hearing aid benefit measured with the PHAB could be predicted on the basis of information that would be available before a hearing aid fitting, namely, adaptation to hearing loss and reported difficulties without a hearing aid.

Adaptation to hearing loss was quantified using Parts II and III of the Communication Profile for the Hearing Impaired, or CPHI (Demorest and Erdman, 1986). The CPHI is a 145-item self-assessment inventory that was designed to quantify communication performance, needs, and strategies, as well as personal adjustment in the hearing-impaired. In the present study, a subset of 127 items was used. Each item was a statement such as "I avoid conversing with others because of my hearing loss." The subject's task was to respond on a five-point scale that indicated either the frequency with which the statement was true (58 items) or the extent of agreement/disagreement with the statement (69 items). All items were scored on a scale of one to five so that a higher score was indicative of fewer problems.

The complete CPHI yields 25 scales. The 127 items used in this study produced 13 scale scores. The 13 scale scores were combined by simple averaging to produce three composite scores, dubbed Communication Strategies, Personal Adjustment, and Problem Denial. The communication strategies (Comstr) score was computed by combining scale scores for maladaptive behaviors, verbal strategies, and nonverbal strategies. The personal adjustment (Peradj) score was produced using scale scores for self-acceptance, acceptance of loss, anger, displacement of responsibility, exaggeration of responsibility, discouragement, stress, and withdrawal. The problem denial (Proden) score was the mean value derived from scale scores for problem awareness and denial. The combinations of scales were selected on the basis of similarity in content as well as statistical similarity reported by Demorest and Erdman (1989).

Subjects

Fifty-eight individuals with essentially sensorineural hearing loss participated in the study. Ages ranged from 38 to 84 with a mean of 68. Seventy-four percent were older than 64 years. Speech reception thresholds ranged from 6 to 70 dB HL (re: ANSI, 1989). The distribution of audiometric data is summarized in Table 1. Most subjects (83%) were mild or moderately hearing impaired with sloping audiograms.

Table 1 Distribution of Audiometric Data for Subjects in Experiment 1

	Slope			
SRT	*<6*	*6–14*	*>14*	*Total*
<40	2	22	26	50
40–60	9	21	14	44
>60	3	3	0	6
Total	14	46	40	100

Data are in percentages and depict functioning of the better ear of each subject.

SRT = speech reception threshold for spondee words (dB HL re: ANSI, 1989); Slope = slope of audiogram from 500 to 4000 Hz in dB/octave.

All subjects wore conventional, analog hearing aids. Half of the fittings were binaural, and 84 percent of them were in-the-ear instruments. The distribution of reported hearing aid experience and daily use is summarized in Table 2. Almost all subjects had worn amplification for more than 1 year and most used their instrument(s) more than 4 hours per day. Six of these subjects had participated in the study reported by Cox et al (1991). A full year elapsed between data collections in the two studies.

Procedure

Paper-and-pencil format was used for both inventories. Most of the subjects completed the PHAB and the CPHI in a single laboratory session. The audiologist explained the instructions, answered questions, and remained available throughout the testing. The PHAB was administered first. The subject then received a basic audiologic evaluation. Finally, the CPHI

Table 2 Distribution of Hearing Aid Experience and Hours of Daily Hearing Aid Use for Subjects in Experiment 1

Hearing Aid Experience	*Daily Use (Hours)*				
	<1	*1–4*	*4–8*	*8–16*	*Total*
6 wk–11 mo	0	2	0	2	4
1–10 yr	5	17	22	28	72
>10 yr	0	2	8	14	24
Total	5	21	30	44	100

Data are in percentages.

was completed. The rest of the subjects completed the CPHI in a clinic environment and were contacted subsequently by mail to complete the PHAB.

RESULTS

For each subject, self-assessed benefit was computed for the seven PHAB subscales. Also, estimated frequency of unaided problems was determined by computing subscale scores for the "without hearing aid" responses to the PHAB. Thus, the PHAB data yielded seven benefit subscale scores and seven corresponding subscale scores reflecting unaided performance. CPHI responses were tallied to produce 13 scale scores for each subject. These scale scores were then combined, as described earlier, into three composite scores.

Figure 1 illustrates the means and standard deviations of subscale scores for the without-hearing-aid condition, compared with those reported by Cox et al (1991). The pattern of scores was similar across the two studies, especially for the five subscales assessing speech communication, FT through BN. The lowest proportion of communication problems was reported for subscales FT and EC, the two subscales relating to face-to-face communication in relatively low-noise environments. The highest proportion of problems was seen in subscales RV and BN, which assess speech communication in reverberant and noisy environments, respectively. Subscale RC, which relates to communication with reduced auditory or visual cues also garnered a fairly high score. The two subscales that relate to perception of environmental sounds, DS and AV, both received rather low scores, indicative of few problems with the quality or aversiveness of sounds.

Figure 2 depicts the means and standard deviations of benefit subscale scores compared with those reported by Cox et al (1991). The dispersion of benefit scores was similar across the two groups. Mean scores were similar, although differences of 7 to 8 percent were seen for FT, DS, and AV. T-tests between mean scores for each subscale failed to reveal any significant differences ($p > .01$) between the two studies.

In the earlier study, significantly more benefit was reported for the FT and EC subscales than for the RC and BN subscales. To explore the pattern of benefit across the speech communication subscales in the present study, a one-way repeated-measures analysis of variance (ANOVA) was performed, followed by a Student-Newman-Keuls post hoc test. The results revealed that subscales FT and RC earned significantly less benefit than subscales EC, BN, and RV ($p < .01$). Thus, the two studies did not result in the same pattern of benefit across speech communication subscales.

Table 3 shows the internal consistency reliability coefficients for the benefit subscales in this study compared with those found in the previous study. For the most part, the coeffi-

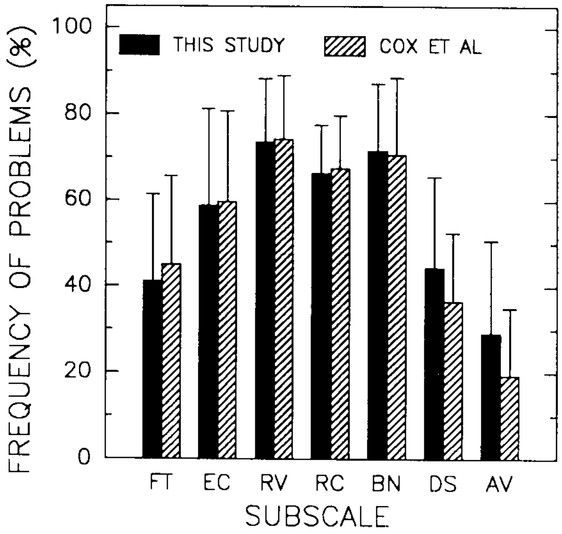

Figure 1 Means and standard deviations of PHAB subscale scores for the without-hearing-aid condition compared with those reported by Cox et al (1991).

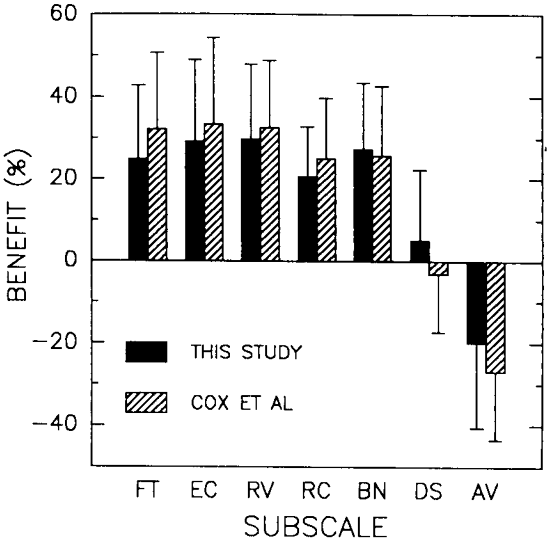

Figure 2 Means and standard deviations of PHAB benefit scores for each subscale compared with those reported by Cox et al (1991).

Table 3 Internal Consistency Reliability (Coeff α) for PHAB Subscales

	C, G, & A	This Study	
Subscale	*Coeff α*	*Coeff α*	*N*
Familiar Talkers (FT)	.88	.85	55
Ease of Communication (EC)	.79	.74	57
Reverberation (RV)	.69	.73	54
Reduced Cues (RC)	.54	.28	55
Background Noise (BN)	.87	.83	56
Distortion (DS)	.38	.20	55
Aversiveness (AV)	.81	.86	54

Data are from the present study (*This Study*) and from Cox, Gilmore, and Alexander (1991) (*C,G, & A*). Sample sizes vary due to missing data on some items.

cients were similar in the two studies. However, subscales RC and DS, which furnished the lowest coefficients in the earlier study, were lower still in the present work.

Means and standard deviations of the 13 CPHI scales are shown in Table 4 and compared to the analogous statistics reported by Demorest and Erdman (1989) for active-duty military service members. Similar norms have been reported for other groups, including elderly persons (Erdman et al, 1990). Overall, the CPHI data obtained in this study were similar to the published norms, suggesting that this group of hearing-impaired persons was quite representative in terms of adaptation to hearing loss.

Means, standard deviations, and intercorrelations for the three CPHI composite scores are shown in Table 5. Note that the correlations between Comstr and the two other scores were low but there was a moderately high negative correlation between Proden and Peradj.

Table 4 Means and Standard Deviations for CPHI Scales

	This Study		D & E	
Scale	*Mean*	*SD*	*Mean*	*SD*
Maladaptive Behaviors	3.90	.65	3.95	.67
Verbal Strategies	2.95	.71	2.97	.76
Nonverbal Strategies	3.93	.63	3.69	.74
Self-Acceptance	3.60	.87	3.50	.91
Acceptance of Loss	3.85	.76	3.70	.76
Anger	3.34	.79	3.27	.78
Displ. of Responsibility	2.88	.69	3.17	.69
Exagg. of Responsibility	2.91	.67	2.97	.76
Discouragement	3.45	.81	3.42	.81
Stress	3.27	.85	3.14	.87
Withdrawal	3.16	.85	3.31	.90
Denial	3.28	.71	3.36	.80
Problem Awareness	4.02	.48	4.02	.55

Data from the present study (*This Study*) are compared to norms from Demorest and Erdman (1989) (*D & E*).

To evaluate the extent to which benefit could be predicted from subscale unaided scores and/or CPHI composite scores, stepwise multiple regression analyses were run. For each subscale, the four potential predictor variables were: subscale unaided score, Peradj, Proden, and Comstr. The outcome indicated that the speech communication subscales yielded different results from the sound perception subscales, as described below.

Speech Communication Subscales

For the five speech communication subscales, the unaided scores made a significant contribution to the prediction of benefit scores, but the addition of CPHI composite scores did not significantly improve the prediction of benefit. Correlation coefficients and regression equations using subscale unaided scores as the only predictor of benefit are shown in Table 6. All five subscales produced significant correlations between unaided scores and benefit ($p < .01$, 1-tailed), although the correlation for subscale RC was marginal. However, the coefficients were rather small except for subscale FT. Thus, only a small proportion of the variance in benefit could be accounted for by the variance in unaided scores. We conclude that knowledge of a patient's unaided scores for the PHAB speech communication subscales would give general guidelines about expected benefit from a hearing aid but precise predictions of benefit would not be possible. This conclusion is bolstered by the relatively large standard errors of estimate shown in Table 6.

Although the regression equations in Table 6 will not produce dependable prediction of benefit in individual cases, they do provide interesting insight into the group patterns of results for these subscales. For all five subscales, the equation constants are small enough to ignore. Thus, we can conceptualize benefit as consisting of a proportion of the unaided score for each subscale. The equations suggest that

Table 5 Descriptive Statistics and Intercorrelations for the Three Composite CPHI Scores

Score	*Mean*	*SD*	*Proden*	*Peradj*
Comstr	3.59	.35	.07	.22
Proden	3.65	.51		−.78
Peradj	3.31	.65		

Comstr = communication strategies; Proden = problem denial; Peradj = personal adjustment.

Table 6 Regression Statistics for the Prediction of Self-Assessed Benefit from Reported Unaided Communication Problems

Subscale	*Equation*	*See*	*r*
FT	.70(UFT)* - 3.8	11.1	.79
EC	.52(UEC) - 1.2	16.3	.58
RV	.40(URV) + 0.5	17.4	.32
RC	.29(URC) + 1.0	11.9	.27
BN	.43(UBN) - 3.2	14.9	.41

*Subscale unaided score.

See = standard error of estimate; r = correlation coefficients between benefit and unaided scores (N = 58).

the typical subject will realize a communication improvement in 70 percent of the FT-type situations (i.e., situations of the type assessed by subscale FT) that are problematic without a hearing aid. Similarly, about 50 percent of the problematic EC-type situations will be alleviated by hearing aid use. Only 30 to 40 percent of troublesome unfavorable listening situations (assessed by RV, RC, and BN) will be improved with amplification. These results mean that, although the magnitude of reported benefit was less for subscale FT than for subscales BN and RV (see Fig. 2), amplification relieved a greater proportion of pre-existing problems in favorable listening situations (assessed by FT and EC) than in unfavorable listening situations (assessed by BN, RC, and RV) for the typical subject.

Sound Perception Subscales

The regression analyses produced a different result for the two sound perception subscales, DS and AV. Of the four predictor variables, the unaided scores were again most closely related to the benefit scores and were entered on the first step of the stepwise analyses. However, in contrast to the speech communication subscale results, a CPHI composite score, Proden, was entered on the second step and made a significant additional contribution to benefit prediction. For subscale DS, 23 percent of the variance in benefit scores was accounted for by unaided scores and 33 percent of the variance in benefit could be accounted for by a combination of unaided score and Proden. For subscale AV, 17 percent of the variance in benefit was accounted for by unaided scores and 29 percent was accounted for by a combination of unaided score and Proden. For both subscales, the relationship between unaided score and benefit was positive, whereas the relationship between Proden and benefit was negative.

EXPERIMENT 2

Reliability of Benefit Scores

In this study, a subset of the subjects from Experiment 1 completed the PHAB on three occasions. These data were used to evaluate the test-retest reliability of the subscale scores and to determine critical differences.

Subjects

Twenty-eight subjects served in this experiment. Distribution of hearing aid experience and use for this subgroup was almost identical to that in Table 2. These subjects were a bit older, on average; 86 percent of them were over 64 years of age. The distribution of audiometric slopes was close to that shown in Table 1 but the distribution of speech reception thresholds revealed a shift towards milder losses, with 75 percent having SRTs lower than 40 dB HL.

Procedure

Originally, all 58 subjects from Experiment 1 were sent a new PHAB inventory by mail about 3 months after the first one was completed. They were informed that the inventory had been slightly changed and they were asked to complete it again if they were still wearing the same hearing aid. It was emphasized that they should complete the inventory without consulting other persons and that they should not attempt to remember their past responses. Instead, they were requested to respond to indicate their current opinions about the instrument.

Thirty-three subjects returned usable inventories. Preliminary statistical analysis (described later) indicated a small but significant change in benefit between the first and second sets of data. To explore this further, a third set of PHABs was mailed to the 33 subjects about 6 months after the second set. Instructions were essentially the same. Twenty-eight usable inventories were returned on this round. These 28 subjects comprised the final group in Experiment 2. The mean time between completion of the first and second inventories was 12 weeks. Twenty-three weeks elapsed, on average, between the second and third PHAB administrations. The shortest inter-test interval for any subject was 6 weeks for the first pair of inventories and 7 weeks for the second pair.

RESULTS

On each measurement occasion, hearing aid benefit was computed for each PHAB subscale. As mentioned above, the initial analysis explored retest differences across the first and second measurements for 33 subjects. A repeated-measures ANOVA with two variables (Test x Subscale) was performed. Results indicated a significant test-retest difference in benefit with the overall benefit of 16.1 percent on the first test significantly less than the overall benefit of 18.8 percent on the second test (F(1,32) = 3.89, p = .05). The interaction was not significant, indicating that this outcome was consistent across subscales. Further exploration of these data revealed that the improved benefit occurred because subjects rated their aided performance as significantly better on the second occasion whereas their estimate of unaided performance remained constant.

After the third measurement occasion, subscale scores were computed for the final 28 subjects. Figure 3 illustrates the means and standard deviations for each subscale. The small improvement in benefit between the first and second administrations can be seen. However, mean subscale scores for the third occasion generally did not maintain this trend. Another ANOVA, using data from all three administrations, indicated that there was no significant effect due to measurement occasion (F(2,54) = 0.56, p > .05). Mean benefit for first, second, and third tests was 16.6 percent, 18.4 percent, and 17.7 percent, respectively. Because the analysis of all three administrations did not reveal a significant difference due to test occasion and because of the small size of the differences between the first and second measurements, we concluded that the mean test-retest differences were either insignificant or of no practical significance.

Correlation coefficients were computed for each subscale between the first and second test (N = 33) and between the second and third test (N = 28). These correlations do not reflect the absolute repeatability of individual subscale scores but they are indicative of the extent to which individuals maintained their relative order across test occasions, that is, whether persons yielding relatively high (or low) scores on one test occasion also yielded relatively high (or low) scores on the next test occasion. For each subscale, the two correlation coefficients were rather similar and they did not show any trend to increase or decrease over time. Table 7 depicts the average coefficient for each subscale. All of the correlations were modest, ranging from 0.42 for subscale DS to 0.72 for subscale AV.

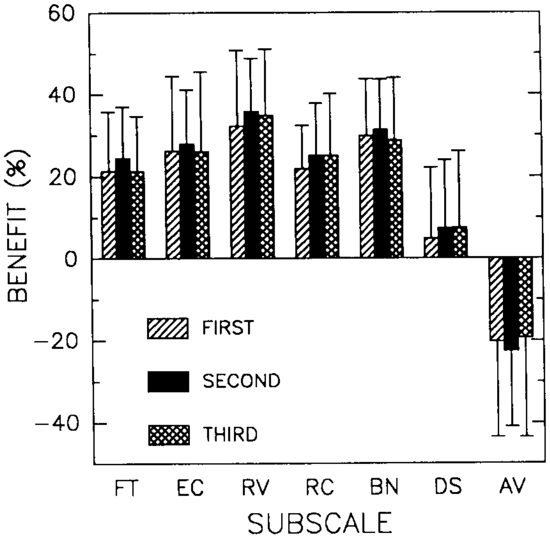

Figure 3 Means and standard deviations of PHAB benefit scores for each subscale on each of three measurement occasions.

The correlation coefficients in Table 7 depict the extent to which inter-subject differences were preserved from test to test for each subscale. It was also of interest to determine the extent to which inter-subscale differences were preserved from test to test for each subject. In other words, we wished to know for each subject whether the shape of the benefit profile obtained on one test occasion was reproduced on the next test occasion. To evaluate this, correlation coefficients were computed between the first and second profiles (N = 33) and between the second and third profiles (N = 28) for each subject (seven pairs of subscale scores per computation). When the shape of the profile remained rather constant across administrations, the correlation coefficient was relatively high. Figure 4 illustrates the distributions of

Table 7 Mean Test-Retest Correlation Coefficient for Each PHAB Subscale

Subscale	*Mean r*
Familiar Talkers (FT)	.50
Ease of Communication (EC)	.54
Reverberation (RV)	.55
Reduced Cues (RC)	.45
Background Noise (BN)	.57
Distortion (DS)	.42
Aversiveness (AV)	.72

r = mean correlation coefficient.

correlation coefficients. Most of the correlations were high or moderate, indicating good reproduction of profiles across test occasions. However, a small proportion of subjects obtained low correlation coefficients revealing that the shape of their benefit profiles varied substantially across test occasions.

The distribution of test-retest differences in benefit was computed for each subscale for the first versus second test (N = 33) and for the second versus third test (N = 28). For each subscale, the standard deviations of these distributions may be used to estimate critical differences (CDs). The CDs, in turn, can be used to evaluate the significance of differences between two scores from the same individual obtained under different conditions (perhaps two different hearing aids). The difference between the two scores will exceed the 95 percent (or 90%) critical difference by chance alone on 5 percent (or 10%) of comparisons. In other words, if the difference between two scores from the same individual exceeds the 95 percent CD, there is only a 5 percent likelihood that this difference was due to chance.

For each subscale, the standard deviations for the two pairs of administrations were combined on a root-mean-square basis to yield the best estimate of the standard deviation. This value was then multiplied by 1.96 and 1.65 to yield the 95 percent and 90 percent CDs, respectively (Table 8). The 90 percent CDs range from 22 percent to 32 percent. The 95 percent CDs range from 25 percent to 38 percent.

DISCUSSION

Comparison with Previous PHAB Data

In most respects, average hearing aid benefit measured with the PHAB was similar in this study to that reported by Cox, Gilmore, and Alexander (1991). Mean and dispersion of scores were similar for all subscales (see Fig. 2) and internal consistency reliability coefficients were similar for five of the seven subscales (see Table 3). Nevertheless, there was a noteworthy discrepancy: the pattern of significant differences across subscales that was observed in the earlier study was not reproduced in this second study. Cox et al (1991) noted that self-report benefit was greater for two subscales that assess favorable listening situations (FT and EC) than for two assessing unfavorable settings (RC and BN). In the present data, FT and RC produced less benefit than EC and BN. There is no obvious explanation for this difference in outcome. The two subject groups had somewhat different hearing loss distributions—subjects in the previous study had a higher proportion of flat and mild hearing losses. However, they produced similar profiles for the without-hearing-aid condition (see Fig. 1) indicating that unaided communication functioning was similar in both groups. Because unaided functioning was equivalent, and because the absolute difference in mean scores for each subscale was small, it might reasonably be argued that this difference in outcome was due to sampling effects. In that case, the best estimate of typical benefit for each subscale would be obtained by combining the two studies. For the interested reader, a family of equal-percentile profiles for successful hearing aid users from both studies is presented in Appendix A, together with a potential application of these profiles.

The small internal consistency reliability coefficients obtained for subscales RC and DS establish even more forcefully than in the previous study that scores for these subscales should not be generalized to situations that are not encompassed by the specific subscale items. It would be defensible to drop these two subscales altogether, thus reducing the inventory to 51 items. However, these subscales may prove

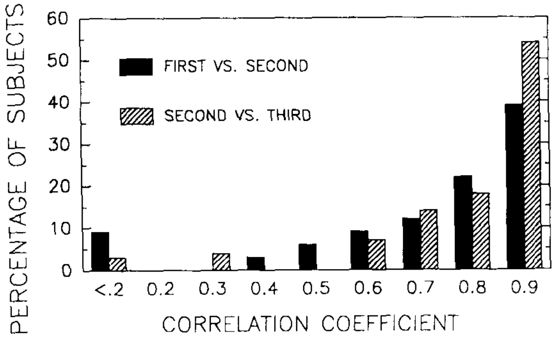

Figure 4 Distribution of correlation coefficients between the first and second profiles and between the second and third profiles.

Table 8 Critical Differences for Each PHAB Subscale

Subscale	*90%*	*95%*
Familiar Talkers (FT)	22	26
Ease of Communication (EC)	27	32
Reverberation (RV)	25	29
Reduced Cues (RC)	23	27
Background Noise (BN)	21	25
Distortion (DS)	32	38
Aversiveness (AV)	27	32

useful despite their limited generalizability. For example, scores on subscale DS were recently found to correlate most highly with results on the Assessment of the Likelihood of Success with Amplification (ALSA), a scale that is completed prospectively by the audiologist to predict hearing aid benefit before a hearing aid is fitted (Cox, 1991).

An intriguing aspect of these results was the relatively large benefit reported by subjects for subscales RV and BN. Because hearing aid wearers often say that they obtain limited benefit in noisy and reverberant situations and most benefit in quiet situations, we expected to observe that benefit was greater in subscales FT and EC than in subscales RV, RC, and BN. Why were these expectations not fulfilled? There are several possible reasons. First, it may be important to keep in mind that the PHAB data do not directly reflect the magnitude of communication improvement as a result of amplification. Instead, they indicate the proportion of situations in which communication is improved. It is possible that hearing aids improve communication slightly in a large number of noisy and reverberant situations and improve it substantially in a smaller number of quiet situations. Alternatively, the comments of amplification wearers may reflect the proportion of their unaided communication problems that are alleviated by hearing aids. The present data indicate that most of the unaided problems that exist in low-noise listening situations are alleviated by amplification whereas a smaller proportion of unaided problems are relieved in noisy and reverberation situations. Finally, we must consider the extent to which the PHAB subscales address the issues to which hearing aid users are responding when they informally report limited benefit in noisy listening situations. Although the content of the items of the PHAB is consistent with that explored by other inventories, its direct relevance to the important experiences of the hearing-impaired may merit further scrutiny.

Predictability of Benefit

It seemed possible that CPHI results for a group of experienced hearing aid users might be different from those reported by Demorest and Erdman (1989) because the latter were obtained principally from subjects who were in the process of receiving their first hearing aid. Presumably, some of Demorest and Erdman's subjects were destined to reject hearing aid use and would not become experienced wearers. As Table 4 shows, however, no such difference was seen in the mean data for the 13 scales used in this study. It appears, therefore, that the typical individual obtaining his or her first hearing aid is similar to the typical experienced hearing aid user in terms of adaptation to hearing loss.

The results of this investigation indicated that adaptation to hearing loss (measured by the CPHI) was not a useful predictor of self-assessed improvement in speech communication due to amplification (as quantified by the five speech communication subscales of the PHAB). On the other hand, self-report of communication difficulties without a hearing aid was a valuable predictor of communication benefit, especially for favorable listening situations. Overall, individuals who reported more difficulty in daily life without a hearing aid were likely to report more benefit from amplification; however, the relationship between unaided communication difficulties and benefit was not strong and a large proportion of variance in benefit was unaccounted for. The CPHI data suggest that this unexplained variance is not substantially due to differences between individuals in psychological-emotional reactions to hearing loss. Other variables that may account for the unexplained variance in benefit include specific characteristics of the hearing aid fitting that might differ across subjects, aspects of personality, and inter-subject differences in auditory processing abilities. Future investigations should explore the effects of these variables on subjective speech communication benefit.

Even though CPHI data were not related to benefit in the speech communication subscales, the composite scores for Proden were found to be useful predictors of results for the sound perception subscales, DS and AV. It should be recalled that the two sound perception subscales are fundamentally different from the speech communication subscales. Subscale DS quantifies the extent to which amplification changes the perceived quality of sounds. As Figures 2 and 3 reveal, the mean score for DS is typically near zero but there is considerable variability across subjects, as shown in the standard deviations. Subscale AV quantifies the extent to which amplification increases the aversiveness of sounds by making them unacceptably loud. The fact that the mean scores on this subscale are invariably negative indicates that the typical hearing aid wearer reports amplified environmental sounds to be more uncomfortably loud than unamplified sounds (thus, in subscale AV, the term "benefit" is something of a misnomer).

For both sound perception subscales, Proden was negatively related to benefit. This means that subjects with lower (or more negative scores) on the PHAB subscales tended to have a higher score on Proden. Proden is a composite of CPHI scales for problem awareness and denial. These scales probe the respondent's awareness of common communication difficulties and emotional reactions to them. A higher score on Proden would be obtained from a subject who is more willing to acknowledge communication problems and who more readily admits to associated negative feelings. The relationship between DS and AV scores and Proden scores leads to the speculation that individuals who are more willing to make negative appraisals of themselves may also be more likely to make negative evaluations of a hearing aid's performance. This finding has implications for the use of subscales DS and AV to evaluate, for example, the appropriateness of a hearing aid's SSPL90 specifications. Our results suggest that the willingness of the subject to make negative self-appraisals should be considered in the interpretation of DS and AV data.

Another aspect of the results for subscale AV is noteworthy. Audiologists who fit hearing aids regularly hear complaints that a hearing aid makes sounds too loud. This is often cited as a primary cause for hearing aid rejection. To protect patients from discomfort, we may measure loudness discomfort levels and adjust hearing aid saturation levels in proportion to them. This practice derives from the assumption that individuals who evidence low tolerance for loudness are the ones most likely to experience problems with excessive loudness of amplified sounds. The present results for subscale AV may not be consistent with this assumption. To be consistent, we would expect those who yield the highest unaided AV scores (indicating low tolerance for the loudness of everyday sounds) to report the greatest aversiveness of amplified sounds. Instead, the positive relationship between unaided AV scores and AV "benefit" indicates that those with the highest unaided AV scores tended to report relatively small increases in loudness discomfort from their hearing aids. Subjects who reported relatively few unaided problems with loudness were more likely to report large increases in negative reactions to amplified sounds. One possible explanation for this outcome is that persons with loudness tolerance problems, evidenced by high unaided AV scores, may be more likely to be wearing hearing aids with reduced saturation levels.

Test-Retest Characteristics of the PHAB

One factor that undoubtedly contributes to the unexplained variance in benefit scores is measurement error. The results of Experiment 2 suggest that measurement error in PHAB scores is not trivial, at least over the relatively long time span covered by this study. The test-retest correlations for each subscale (see Table 7) revealed that, except for AV, the correlations were modest. Only about 20 to 30 percent of the variance in benefit scores on the second measurement could be accounted for by the variance in benefit scores on the first measurement. Because the subjects were all experienced hearing aid wearers and were all using the same hearing aid throughout the study, it is unlikely that their opinions about hearing aid benefit underwent a true change during this time. Therefore, it is probable that test-retest differences were due in part to day-to-day variations in mood, health, or recent experiences.

Limited comprehension of items may also have contributed to retest variability. Even more modest retest correlations, ranging from 0.28 to 0.42, were reported by Demorest and Erdman (1988) for the Communication Performance scales of the CPHI. These scales do not measure hearing aid benefit, but are similar to the speech communication subscales of the PHAB because they quantify communication in daily life. Demorest and Erdman (1988) suggested that their subjects may have had difficulty understanding the instructions for the communication items. A report by Weinstein, Spitzer, and Ventry (1986) also supports this notion. They found that a face-to-face administration method produced more reliable self-report scores for the Hearing Handicap Inventory for the Elderly than a paper-and-pencil administration. Problems of comprehension may be especially relevant for the PHAB because the items are written so that a response of "always" sometimes indicates severe problems and sometimes indicates few problems. Thus, subjects must pay close attention to the wording of each item as well as its content area. It seems possible that vigilance in this task may lapse from time to time, reducing the reliability of the data.

If a subject fails altogether to notice the details of item wording, this tends to produce a characteristic result, namely, a large negative "benefit" for subscale EC. Such a result suggests that communication becomes substantially more difficult when a hearing aid is worn. We consider this to be, *ipso facto*, evidence of an invalid response pattern. Because a small proportion of subjects have shown this pattern, it is recommended to supplement the written instructions of the PHAB with a specific reminder to "pay close attention to the words of each item because this can have an important effect on the meaning."

It was encouraging to note that the shape of the benefit profiles was fairly constant across measurement occasions for a large proportion of subjects (see Fig. 4). Because the principal advantage of the PHAB as a measure of benefit is its promise of an analytical evaluation, this perspective on data reliability may produce information that is of greater practical significance than test-retest correlations for individual subscales. To illustrate, if a subject produces a high benefit score for subscale FT in association with a low benefit score for subscale BN, we would probably interpret this as indicative of a need for amplification adjustment to improve intelligibility of noisy speech. However, we need to be reasonably confident that this response pattern is reproducible before undertaking a program of intervention. The data shown in Figure 4 suggest that, for most subjects, profile patterns have either moderate or excellent reproducibility.

Test-retest correlation coefficients furnish insights about the extent to which the pattern of differences between subjects, or subscales, is preserved from one test to the next. However, they do not facilitate judgments about the absolute consistency of an individual's scores over test occasions. When we need to know whether the scores for an individual are significantly more or less on a second test occasion, the difference between scores can be evaluated by comparison with a critical difference such as those reported in Table 8. The 95 percent critical differences for the subscales of the PHAB ranged from 25 percent to 38 percent. This means, for example, that two benefit scores from the same individual for subscale FT must differ by 26 percent or more before we can conclude with reasonable certainty that there is a real difference between them. These CDs are 5 to 10 percent larger than CDs for the same subscales measured using the Profile of Hearing Aid Performance (PHAP), the inventory on which the PHAB is based (Cox and Gilmore, 1990). The difference in CD size between the PHAP and the PHAB is not surprising when we consider that the PHAP quantifies only aided performance whereas the PHAB must quantify both aided and unaided performance, requiring twice as many subjective judgments per inventory.

Although no previous studies have specifically addressed the absolute consistency of subjective hearing aid benefit data, other studies of self-report measures have produced comparable CDs for subjective data from hearing-impaired persons. Weinstein et al (1986) reported a 95 percent CD of 36 percent for the pencil-and-paper version of the Hearing Handicap Inventory for the Elderly. Demorest and Erdman (1988) reported the distribution of test-retest differences for the 25 scales of the CPHI. When the average test-retest SD was used to generate a 95 percent critical difference analogous to those reported here, the CD was found to be 1.21 scale intervals or 30 percent of the scoring range. These values are similar to those found in the present study.

We conclude that the absolute consistencies of hearing aid benefit subscale scores from the PHAB are typical of those expected using paper-and-pencil inventories with hearing-impaired individuals. On the other hand, the size of the critical differences is quite large compared to the expected size of treatment effects that might be assessed using this type of measurement vehicle. For example, one possible application for a self-report benefit inventory would be to evaluate the differences in daily life between two opposing amplification strategies (perhaps linear versus compression amplification). The present data indicate that, if data are to be evaluated on an individual basis, hearing aid fittings implementing the two strategies would have to produce about a 30 percent difference in benefit in at least one subscale to be judged truly different. A benefit difference of this size seems unlikely except under extraordinary circumstances. Furthermore, true benefit differences considerably smaller than 30 percent would probably be important.

This outcome suggests that a single administration of the PHAB under each condition often would not be sensitive enough to detect important differences between conditions for a single individual. Sensitivity could be improved by administering the PHAB several times in each condition and averaging the results. If this were done, it would be necessary to take steps to minimize the likelihood that subjects would remember previous responses. An alternative approach might be to use a single-subject type of research design in which test conditions are alternated several times and the judgment of a difference between them is based on the pattern of scores rather than a critical difference value. Also, because the PHAP has smaller critical differences than the PHAB, it would be advisable to use the PHAP in studies where an estimate of unaided performance is not necessary. These caveats apply only for applications of the PHAB where it is desired to make decisions about treatment effects for specific individuals. The limitations are much less when group statistics can be used to judge the significance of score differences.

Finally, it should be noted that certain aspects of this investigation might have negatively influenced the reliability of PHAB subscale scores. These include the rather long time interval between tests and the administration of tests under relatively uncontrolled home conditions. It is possible that a test-retest time span of, say, 2 to 3 weeks, together with laboratory administration of the inventory on each occasion, would produce more reliable data.

CONCLUSIONS

This study demonstrated that some of the between-subject differences in self-assessed hearing aid benefit can be accounted for by reported amount of unaided difficulties and, to a more limited extent, by the individual's willingness to make negative self-appraisals. Nevertheless, the largest portion of benefit variability between subjects could not be accounted for in this way. One of the major contributors to inaccurate benefit predictions was probably the limited reliability of the benefit measures themselves. Future studies should explore other variables that may be related to intersubject differences in hearing aid benefit and should attempt to determine the conditions necessary to maximize the reliability of self-report benefit data.

Acknowledgment. This work was supported by the Department of Veterans Affairs Rehabilitation Research and Development Service. Suzanne House assisted with data collection.

REFERENCES

American National Standards Institute. (1989) *American National Standard Specification for Audiometers*. (ANSI S3.6-1989). New York: ANSI.

Brooks DN. (1989) The effect of attitude on benefit obtained from hearing aids. *Br J Audiol* 23:3–11.

Cox RM. (1991) *Studies of Hearing Aid Benefit*. Presented at the National Convention of the American Academy of Audiology, Denver, CO.

Cox RM, Gilmore C. (1990) Development of the Profile of Hearing Aid Performance (PHAP). *J Speech Hear Res* 33:343–357.

Cox RM, Gilmore C, Alexander GC. (1991) Comparison of two questionnaires for patient-assessed hearing aid benefit. *J Am Acad Audiol* 2:134–145.

Demorest ME, Erdman SA. (1986) Scale composition and item analysis of the Communication Profile for the Hearing Impaired. *J Speech Hear Res* 29:515–535.

Demorest ME, Erdman SA. (1988) Retest stability of the Communication Profile for the Hearing Impaired. *Ear Hear* 9:237–242.

Demorest ME, Erdman SA. (1989) Factor structure of the Communication Profile for the Hearing Impaired. *J Speech Hear Disord* 54:541–549.

Erdman SA, Demorest ME, Wark DJ, Lansing CR, Henoch MA, Corn L, Binzer SM. (1990) *Factors Affecting Adjustment to Hearing Loss*. Presented at the National Convention of the American Speech-Language-Hearing Association, Seattle, WA.

Green R, Day S, Bamford J. (1989) A comparative evaluation of four hearing aid selection procedures. II: Quality judgments as measures of benefit. *Br J Audiol* 23:201–206.

McClymont LG, Browning GG, Gatehouse S. (1991) Reliability of patient choice between hearing aid systems. *Br J Audiol* 25:35–39.

Weinstein BE, Spitzer JB, Ventry IM. (1986) Test-retest reliability of the Hearing Handicap Inventory for the Elderly. *Ear Hear* 7:295–299.

APPENDIX A

PHAB Results from Successful Hearing Aid Wearers

Figure A1 illustrates a family of equal-percentile profiles from 64 successful hearing aid wearers, derived by combining subjects from this study and from Cox et al (1991). An individual was considered a successful user if he or she had worn the hearing aid for more than 1 year and reported a daily use of 4 or more hours.

This type of display may be useful in evaluating the results for a particular individual and may facilitate identification of specific problem areas. To illustrate, PHAB results are depicted for two individuals, labelled "A" and "B." Although both subjects had owned their hearing aids for more than 1 year, their reported use was less than 1 hour per day. Thus, they were judged to be unsuccessful hearing aid wearers.

As Figure A1 shows, subject A reported rather good benefit for all of the speech communication subscales of the PHAB: His scores are in the 65th to 80th percentile range. However, for the sound perception subscales, his scores drop into the 5th to 35th percentile region, indicating that, in terms of quality and aversiveness of amplified sounds, his scores were relatively poor. These results suggest that the negative aspects of amplified sounds may be the main reason for this subject's poor adjustment to hearing aid use. Investigation of the distortion properties or maximum output of the instrument would seem to be indicated.

Subject B, on the other hand, appears to have a different problem. On the sound perception subscales, this individual's scores were in the 50th to 65th percentile region, indicating that his reaction to the quality and aversiveness of amplified sounds was better than the median for successful wearers. However, he reported essentially zero benefit for the speech communication subscales. In this case, a general lack of communication benefit might well account for poor adjustment to amplification. Perhaps this subject's needs could be met more satisfactorily with an entirely different instrument or with alternative assistive listening devices.

Although these speculations seem plausible, the actual utility of this approach to evaluation of unsatisfactory hearing aid fittings can be established only through empirical tests. Research is needed to determine whether comparison of data from unsuccessful and successful hearing aid wearers can indicate the appropriate direction for intervention.

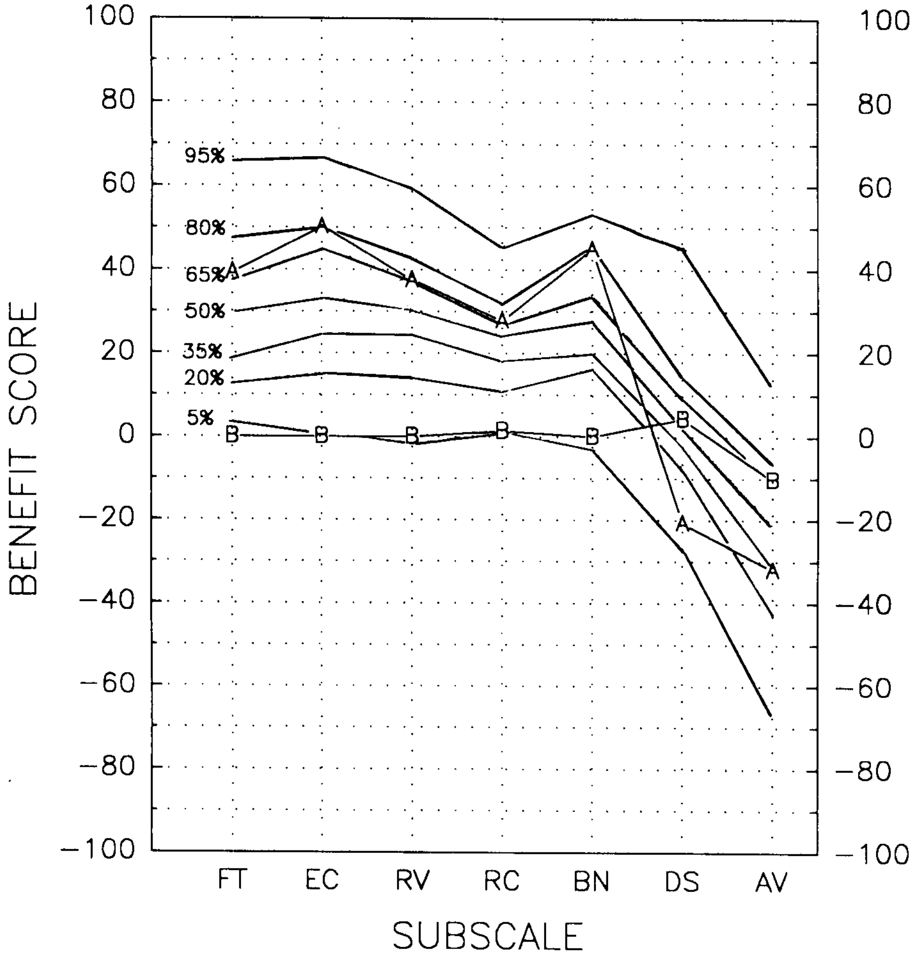

Figure A1 A family of equal-percentile PHAB profiles for successful hearing aid wearers. Each solid line without symbols depicts the benefit score for each subscale at a particular percentile level. For example, 20 percent of successful hearing aid wearers give subscale scores that are less than the line labelled "20%" (80% of successful wearers give subscale scores that exceed this line). The figure also shows PHAB profiles for two unsuccessful hearing aid wearers, represented by *A* and *B*.