Self-Evaluation of Voice as a Treatment Outcome Measure

Anne-Maria Laukkanen  Kirsti Leppänen  Irma Ilomäki
Department of Speech Communication and Voice Research, University of Tampere, Tampere, Finland

Key Words
Voice quality • Vocal fatigue • Voice training • Voice Massage™

Abstract
This study addressed two self-evaluation questionnaires in investigating the effects of voice hygiene lecture (VHL, 3 h) and additional voice training (VT) or Voice Massage™ (VM; both 5 h) in 90 female teachers. The subjects assessed their voice quality, ease of phonation and tiredness of throat before and after a working day at the beginning and end of the school term using a visual analogue scale (VAS) (Questionnaire 1). At the end of the term, the degree of positive influence of the interventions was reported on VAS, and the type of influence was indicated by choosing one or more of three alternatives (voice quality, audibility and endurance) or by writing a free comment (Questionnaire 2). Questionnaire 1 was pretested and found to be reliable and valid for self-evaluation of voice. At the end of term, the VM and VT groups reported more positive influence of the interventions than did the VHL group. The reported influence did not correlate with working-day-related changes in sensations. Increased difficulty of phonation and tiredness of throat was found in the VHL group at the end of the term. However, the groups did not differ significantly from each other. The challenges of self-evaluations as outcome measures are discussed.

Introduction
Voice training (VT) and therapy aim at preventing and curing voice problems by changing vocal habits. Various studies have shown positive acoustic and perceptual effects of therapy and training [1–4] in terms of decreased perturbation (jitter and shimmer), increased signal-to-noise ratio, decreased spectral tilt, a formation of a sound energy concentration called a speaker’s formant and a better, more resonating voice quality. Sometimes a better voice quality may be manifested in increased spectral tilt and a more relaxed voice production [5]. Voice hygiene education alone has also been reported to have a positive impact [6]. Even better results have been reported after voice hygiene compared to a training period [7]. Various massage methods have been applied to improve voice production by reducing excessive tension in the laryngeal muscles [8, 9]. Some studies have shown beneficial effects of these treatments on dysphonic patients’ voices [10–12]. Ternström et al. [13] investigated the immediate effects of a 30-min (naprapathic) massage of the muscles of the back, neck and face on nor-

This paper is an updated version of a lecture prepared for the special symposium of the IALP Voice Committee at the IALP World Congress in Copenhagen, Denmark, 2007 (Guest Editor: Jan Svec, Groningen).

Anne-Maria Laukkanen, PhD
Department of Speech Communication and Voice Research, University of Tampere
FI-33014 Tampere (Finland)
Tel. +358 3 3551 7249, Fax +358 3 3551 6063
E-Mail Anne-Maria.Laukkanen@uta.fi; www.uta.fi/laitokset/puheoppi
mal-voiced subjects. Naprapathy refers to manual therapy that focuses on the evaluation and treatment of neuromusculoskeletal conditions [14]. According to the results by Ternström et al., the relaxing effect on the muscles involved in voice production was reflected in lower F0 and SPL of text reading after the treatment session. Voice Massage™ (VM) is a Finnish method, developed by massage therapist Leena Koskinen (see www.voicemassage.fi). It consists of manipulation of voice and speech production muscles of the larynx, respiration and articulation. It also includes some vocal and respiratory exercises during manipulation. VM is used for voice students, voice professionals and also for voice patients (especially those with hyperfunctional dysphonia) as a supportive treatment related to logopedic therapy. Laukkanen et al. [15] studied the effects of 1-hour VM treatment on normal-voiced subjects. Improved ease of voice production and relaxation of the neck and back were reported immediately after the treatment. However, no significant acoustic or perceptual changes were found in the speech samples recorded.

The problem in using acoustic variables as outcome measures of training and therapy lies in the facts that opposite acoustic changes (for instance increased and decreased spectral tilt) may be positive depending on the starting point (i.e. if the voice production is hyper- or hypofunctional to start with), and that similar acoustic changes (like decreased spectral tilt) may be due to different causes (increased subglottic pressure and laryngeal resistance or increased vocal tract impedance). Furthermore, no clear relation exists between perceptual voice quality and the laryngeal state, e.g. pathological tissue changes of the vocal folds. Subjective sensations mainly determine whether a person regards him-/herself as being in need of therapy or training. The truthfulness and meaning of subjective ratings is difficult or even impossible to deny. Sometimes, positive sensations are the only outcome obtained in a treatment, as shown in the study by Laukkanen et al. [15].

Various self-evaluation questionnaires have been developed to be used in voice patients to grade the severity of the problems and to serve as the outcome measure of therapy. These questionnaires like Voice Handicap Index (VHI) [16, 17], Voice-Related Quality of Life (V-RQOL) [18–20], Voice Activity and Participation Profile (VAPP) [21], Reflux Symptom Index (RSI) [22], Voice Symptom Severity Index (VoiSS) [23] and Vocal Tract Discomfort Scale (VTD) [12] focus on the type and amount of vocal symptoms and/or their effects on the quality of life. Such questionnaires, however, do not appear to be well suited to be used in voice professionals with functionally healthy voices. The present study tests two simple, easy-to-use questionnaires in (1) disclosing the effects of vocal loading (e.g. related to daily work) and (2) the outcome of various voice hygienic interventions. The characteristics concerned were vocal endurance (i.e. lack or low prevalence of symptoms of vocal fatigue in relation to vocal loading at work), ease of voice production, self-rated (assumed) audibility and voice quality. These characteristics are supposed to be improved through voice hygiene interventions.

Three interventions that aimed to support and improve voice professionals’ vocal working capacity and well-being at work were tested. The interventions chosen were a voice hygiene lecture (VHL), VT and VM. This study focused on teachers, since they are one of the largest professional groups with voice as the main tool of trade, voice problems of teachers are known to be common worldwide, and teachers also form the majority of phoniatric patients [24, 25]. Furthermore, since the majority of the teachers are females and since voice problems are twice as common in females than in males, this study focused on female teachers. Part of the results have been reported in the Proceedings of IALP 2007 [26].

Material and Methods

Ninety Finnish female primary school teachers (mean age 41.1 years, SD 8.5 years; mean experience in teaching 15.3 years, SD 8.8 years; mean teaching hours per week 24.3 h, SD 4.4 h) were recruited for the study via a questionnaire on the internet. The questionnaire concerned age, general health, vocal training, working experience, classroom size, and symptoms of vocal fatigue. All subjects were given a 3-hour VHL. In addition to the lecture, a randomly chosen group of 30 subjects received VM treatment (VM group) and another group of 30 subjects received VT (VT group). The remaining group (n = 30) which received only the VHL constituted the VHL group. Both VT and VM were given in 5 sessions, each lasting 1 h. The first 3 sessions of VM and VT were given with an interval of 1 week, the last 2 sessions with an interval of 4 weeks. Five sessions were chosen since it is the number of therapy sessions generally given to voice patients in Finland, and it is also the traditional number of VM treatments given in succession. A phoniatric inspection was carried out at the beginning of the term to assess the laryngeal status of all participants.

At the beginning and at the end of the term, before and after a vocally loading working day, the subjects rated the ease or difficulty of phonation, tiredness of throat and voice quality using a visual analogue scale (VAS; Questionnaire 1, see Appendix 1). In the case of tiredness of voice, a unipolar 100-mm scale was used, while for the other two parameters a 200-mm bipolar scale was
used, in order to allow the subjects to report about possible positive warm-up effects. At the end of the term, the degree of positive influence of the interventions was reported using a 100-mm VAS (Questionnaire 2, see Appendix 2). The type of influence was reported by choosing one or more of the alternatives given (voice quality, auditory and endurance) or by writing an additional comment.

**Statistical Treatment**

Values obtained before and after a working day and at the beginning and at the end of the term were compared using the paired Student t test. At the end of the term, the three intervention groups were compared to each other using a variance analysis (general linear model, univariate analysis). The values obtained at the beginning of the term were set as covariates. Pearson correlations were used to study the relationship between ratings of the general beneficial influence of the interventions and the changes in working-day-related sensations of voice production and self-evaluations of voice quality. Statistical analyses were carried out with SPSS 15 (SPSS Inc., Chicago, Ill., USA).

### Table 1. Repeated self-assessment of voice on VAS (mm): Questionnaire 1

<table>
<thead>
<tr>
<th></th>
<th>First time</th>
<th></th>
<th>Second time</th>
<th></th>
<th>Diff. between 1st and 2nd time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>before</td>
<td>after work</td>
<td>diff.</td>
<td>before</td>
<td>after work</td>
</tr>
<tr>
<td>Production</td>
<td>92.2 (27.3)</td>
<td>105.7 (22.2)</td>
<td>p = 0.002</td>
<td>89.0 (30.1)</td>
<td>116.3 (37.4)</td>
</tr>
<tr>
<td>Quality</td>
<td>93.6 (25.7)</td>
<td>110.3 (28.3)</td>
<td>p = 0.035</td>
<td>85.6 (36.1)</td>
<td>118.7 (28.6)</td>
</tr>
<tr>
<td>Throat</td>
<td>30.7 (20.1)</td>
<td>57.6 (22.5)</td>
<td>p = 0.000</td>
<td>30.2 (26.5)</td>
<td>54.9 (20.7)</td>
</tr>
</tbody>
</table>

Production: 0 = easy, 100 mm = ordinary, 200 mm = very difficult; quality: 0 = good, 100 = ordinary, 200 = very poor; throat: 0 = not tired, 100 = very tired. Significance of differences: paired Student’s t test [nonsignificant (NS), p ≥ 0.05]. Test repeated at intervals of 1–2 weeks (n = 16 female teachers).

### Table 2. Results of validity test of Questionnaire 1

<table>
<thead>
<tr>
<th></th>
<th>Production</th>
<th>Quality</th>
<th>Tiredness of throat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity score</td>
<td>r = 0.31, p = 0.003</td>
<td>r = 0.21, p = 0.02</td>
<td>NS</td>
</tr>
<tr>
<td>Job score</td>
<td>r = 0.38, p = 0.000</td>
<td>r = 0.30, p = 0.004</td>
<td>r = 0.21, p = 0.049</td>
</tr>
<tr>
<td>Communication score</td>
<td>r = 0.24, p = 0.027</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Social score</td>
<td>NS</td>
<td>NS</td>
<td>r = 0.23, p = 0.034</td>
</tr>
<tr>
<td>Emotion score</td>
<td>r = 0.29, p = 0.007</td>
<td>r = 0.24, p = 0.024</td>
<td>NS</td>
</tr>
<tr>
<td>Total score</td>
<td>r = 0.31, p = 0.005</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

Correlations between self-evaluation of ease/difficulty of voice production (= voice production), voice quality and tiredness of throat and sum scores for each of the five sections of VAPP: (1) self-perceived voice problem score (severity score), (2) effect of voice problem on job (job score), (3) effect on daily communication (communication score), (4) effect on social communication (social score) and (5) effect on emotion (emotion score) and with sum of the five sections (total score). r = Correlation coefficient; NS = nonsignificant (p > 0.05).

**Pretests of Questionnaire 1**

Questionnaire 1 was tested for reproducibility, internal consistency and validity. The results of this pretesting suggest that the questionnaire is suitable for use as an outcome measure of voice hygiene interventions.

**Reproducibility**

Sixteen teachers (not included in the present study) tested questionnaire 1. On two occasions, they evaluated their voices before and after a vocally loading working day. The interval between the evaluations was 1–2 weeks. This has been regarded as an appropriate interval, long enough for the subjects to forget their previous answers and short enough not to permit any substantial changes in their behavior due to learning new vocal habits [18, 19]. Paired t tests, repeated measures analysis of variance (RM ANOVA) and the Pearson correlation coefficient were used in the analyses. After a working day a significant deterioration was seen in the parameters (paired Student’s t test), while no significant differences were found between test and retest (table 1). The effect of repeating the test (RM ANOVA, within subjects’ effects) was nonsignificant, while the effect of time of the day (be-
fore/after working day; between subjects effects) was significant (p = 0.034 for production, p = 0.006 for quality and p = 0.003 for tiredness of throat). The results obtained with Questionnaire 1, thus, seemed to be reproducible.

Internal Consistency

Relations between the three variables (ease/difficulty of voice production, voice quality and tiredness of throat) were studied by calculating Cronbach’s alpha coefficient and Pearson correlation coefficients for the answers of all 90 subjects at the beginning of the term. Internal consistency between the variables was high (alpha = 0.81). Ease/difficulty of voice production correlated strongly with voice quality (r = 0.80, p = 0.000) and moderately with tiredness of throat (r = 0.55, p = 0.000). Tiredness of throat and voice quality also correlated moderately (r = 0.40, p = 0.000).

Validity

Results of the three variables (ease/difficulty of voice production, voice quality and tiredness of throat) of Questionnaire 1 obtained after the working day at the beginning of the term were compared to (1) results obtained from the VAPP questionnaire [21] and (2) results of perceptual voice analysis performed by three experienced speech trainers from 1-min text reading samples recorded after the working day. VAPP describes the perception of voice problems and their impact on a person’s daily activities and social function [21]. VAPP consists of 28 questions, concerning five areas: self-perceived voice problem (1 question), job (4 questions), daily communication (12 questions), social communication (4 questions) and emotion (7 questions). The maximum total score is 280. The questions are answered using a 10-cm VAS, where the left end represents ‘not affected’ and the right end ‘always affected’. The Finnish translation of VAPP, which was used in the present study, has been validated [27].

Samples were recorded using a portable DAT recorder (Sony TCD-D8) and a headset microphone (AKG B29L; mouth-to-microphone distance 6 cm). Perceptual analysis was carried out using Judge software (Svante Granqvist) implemented in Sound-Swell Signal Workstation. Headphones (Sennheiser HD 530 II) were used in the listening evaluation. The following characteristics were evaluated using a 1,000-unit VAS: general voice quality, tightness of voice production (from hypo- to hyperfunctional), breathiness (presence of turbulence noise) and hoarseness. Results of the validity analyses can be seen in table 2. All three variables correlated significantly with many of the sum variables of VAPP, and all variables correlated with the work-related sum (effect of voice problem on job). In contrast, no significant correlations were found between self-reported variables and results of the perceptual analysis. These results show that results obtained with Questionnaire 1 are in accordance with self-evaluation of voice obtained using a standardized questionnaire and may, thus, be seen as valid, based on the generally accepted way of solving criterion validity used in psychometrics.

Results

Positive influence was reported after all interventions, but significantly more (p = 0.000) after VM (mean 66.6 mm VAS, SD 25.9) and VT (mean 56.2 mm VAS, SD 24.7) than after VHL (mean 27.5 mm VAS, SD 19.5). Figure 1 shows the distribution of the evaluations in each intervention group. The positive effects reported mainly con-
Fig. 2. Distribution of self-ratings for ease or difficulty of voice production (a), tiredness of throat (b) and voice quality before and after a working day in the beginning and end of the term (c). Ratings given on a VAS. a Production: 0 = easy, 100 mm = ordinary, 200 mm = very difficult. b Throat: 0 = not tired, 100 = very tired. c Quality: 0 = good, 100 = ordinary, 200 = very poor. Significance of differences (paired Student’s t test): * p < 0.05 (above the bars) = difference before/after a working day; * p < 0.05 (below the bars) = difference between the beginning/end of term.
sisted of improved knowledge in the VHL group, increased relaxation and vocal endurance in VM group and increased knowledge and vocal endurance in the VT group (table 3).

Table 3. Self-reported improvement of voice (Questionnaire 2)

<table>
<thead>
<tr>
<th></th>
<th>Quality</th>
<th>Audibility</th>
<th>Endurance</th>
<th>Knowledge</th>
<th>Relaxation</th>
<th>Body awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHL group</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VT group</td>
<td>10</td>
<td>8</td>
<td>13</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM group</td>
<td>17</td>
<td>8</td>
<td>20</td>
<td>26</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Correlation matrix between results from Questionnaire 1 (self-reported voice quality, ease or difficulty of voice production and tiredness of throat) and Questionnaire 2 (degree of positive influence of the interventions reported in mm VAS)

<table>
<thead>
<tr>
<th>Degree of positive influence</th>
<th>Quality</th>
<th>Production</th>
<th>Tiredness of throat</th>
</tr>
</thead>
<tbody>
<tr>
<td>VHL group</td>
<td>$r = -0.12$, NS</td>
<td>$r = -0.01$, NS</td>
<td>$r = -0.01$, NS</td>
</tr>
<tr>
<td>VT group</td>
<td>$r = -0.12$, NS</td>
<td>$r = -0.06$, NS</td>
<td>$r = -0.004$, NS</td>
</tr>
<tr>
<td>VM group</td>
<td>$r = 0.06$, NS</td>
<td>$r = 0.22$, NS</td>
<td>$r = 0.32$, NS</td>
</tr>
</tbody>
</table>

Pearson correlation coefficient. NS = Nonsignificant.

There was no correlation between the general assessment of the positive influence of the interventions (Questionnaire 2) and the working-day-related changes in sensations (Questionnaire 1) (see table 4).

Discussion

Positive effects were reported after all three interventions. The involvement of psychological effects cannot be excluded. People like it if attention is paid to them, and if they are touched as during a massage. They may also be willing to please and give a positive response when having been subjected to something which is supposed to be beneficial for them. On the one hand, the inclusion of a placebo treatment in the research setup might be worth while. On the other hand, if a person feels pleasure for whatever (nonharmful) reason, this is also likely to have positive physiological effects and may, thus, have true positive effects on vocal function as well.

General assessment of the positive influence of the interventions (Questionnaire 2) did not correlate with working-day-related changes in sensations (Questionnaire 1). This is understandable, since the general assessment considered whatever positive effects there were and these effects may also have been viewed using a different ‘time window’ than 1 working day. On the other hand, a working-day (or in general, vocal loading)-related questionnaire might be more useful to show what practical consequences various treatments may have in voice professionals’ real-life situations.

Increased working-day-related difficulty of phonation and tiredness of throat in VHL group at the end of the term seems to suggest that VM and VT may have helped the teachers to sustain better their vocal well-being during the term. Thus even a short-term (5 sessions) treatment may have a positive effect on vocal well-being. A one day course has also been reported to have beneficial influence on voice professionals [28–30]. However, the three intervention groups did not differ significantly from each other in terms of subjective evaluation at the end of the term when the starting point had been taken into account. Should this be interpreted to show that no real changes had taken place in the intervention groups? At least, it suggests that it is difficult to show any clear changes. There may be many reasons for this. It is possible that the intervention time was too short to result in substantial systematic changes. The ratings of the general positive influence also showed a relatively large dis-
tribution (fig. 1) which may reflect individual differences in learning strategies or psychosocial factors (positive or negative opinion of the lecturer, voice trainer or massage therapist). It is also possible that the relatively low ratings given to the effects of VHL may reflect some subjects’ disappointment with the fact that they were not chosen to either of the other treatment groups. It is also possible that positive changes take place due to mere adaptation to vocal loading during the term and, thus, they may be obtained in the lecture group as much as in the treatment groups. One problem related to self-evaluation as an outcome measure of voice hygiene interventions lies also in the fact that a possible result of the interventions is increased sensitivity to symptoms of vocal fatigue. Finally, it is difficult to control the baseline setting in the type of a questionnaire used for tracking changes during a working day (Questionnaire 1). It might be argued that to some extent the validity of the procedure might be improved by allowing the subjects to see their earlier replies (e.g. those given before a working day) and by repeating the rating procedure on different days and by taking an average of the successive ratings. On the other hand, seeing one’s own earlier results might also bias the responses. In any case, due to limitations of the human memory buffer (when true multisensory experiences are in question), the questionnaire may be more suited to show the immediate effects of for instance vocal loading but may be less able to indicate the possible changes in endurance after a longer-term phenomenon like training.

The acoustic and perceptual effects of these interventions will be investigated in another study. The duration of the effects will also be addressed in two follow-up studies to be conducted 6 and 12 months’ after termination of the project. The need of finding new objective methods for detecting the physiological effects of loading is obvious.

**Summary and Conclusions**

1. Two self-evaluation questionnaires were used as an outcome measure of three voice hygiene interventions: VHL, VM and VT. Questionnaire 1 aimed to measure vocal endurance during a working day. Questionnaire 2 collected opinions of the general effects of the interventions. Questionnaire 1 was tested and found reproducible, reliable and valid as tested against Voice Activity and Participation Profile questionnaire (psychometric criterion validity based on correlation between a new scale and an earlier tested one).

2. Positive effects were reported after all interventions, but significantly more after VT and VM than VHL (Questionnaire 2).

3. Increased difficulty of phonation and tiredness of throat in VHL group at the end of the term may suggest that VT and VM assisted in sustaining the teachers’ vocal well-being during the term.

4. Self-evaluation results before/after working day were not able to show significant effects of the interventions.

5. Shortcomings of self-evaluations should be taken into account when using them as an outcome measure.

6. Reliability of the research setup might be improved by (1) including placebo treatment, (2) using multiple questionnaires and (3) repeating the same rating various times before and after an intervention and at follow-up.

7. A need for reliable measures of physiological vocal loading effects is obvious.

**Acknowledgements**

The present study was supported by the Finnish Work Environment Fund (grant No. 103309), the Scientific Foundation of the City of Tampere and the Voice Massage Therapist Association in Finland.
Appendix 1: Questionnaire 1 for Teachers

Self-assessment of voice before a working day

Background information

Name: ____________________________

Recording place: ________________ Date: ______ Time: ______

How long have you been awake before the morning recording?

Have you used your voice before the morning recording? Yes/No

How have you used your voice before the morning recording?

Please draw a Vertical Line ( ) that describes best your self-perception of your voice in this moment

(1) Voice production in mm VAS 200
   very easy                            ordinary                       very difficult

(2) Voice quality in mm VAS 200
   good                               ordinary                                very poor

(3) Throat in mm VAS 100
   not tired very tired

Self-assessment of voice after a working day

Recording place: ________________ Date: ______ Time: ______

Please draw a Vertical Line ( ) that describes best your self-perception of your voice in this moment

(1) Voice production in mm VAS 200
   very easy                            ordinary                       very difficult

(2) Voice quality in mm VAS 200
   good                               ordinary                                very poor

(3) Throat in mm VAS 100
   not tired very tired

References