Received: 28 March 2023 | Revised: 18 June 2023 | Accepted: 8 July 2023

DOI: 10.1002/lio2.1117

ORIGINAL RESEARCH

Laryngoscope Investigative Otolaryngology

# Mandarin Chinese version of the Aging Voice Index

Chia-En Liao MS[1] | Roger W. Chan PhD[2,3] | Peiyun Zhuang MD[3] | Chi-Te Wang MD, PhD[4] | Feng-Chuan Lin MS[4] | Ying Chen PhD[2] | Huiqin Yu MS[2]

[1]Sure Care Rehabilitation Clinic, Taipei, Taiwan

[2]Geriatric Care Research Center, Xiamen Medical College, Xiamen, Fujian, China

[3]Department of Voice Medicine, Zhongshan Hospital Xiamen University, Xiamen, Fujian, China

[4]Department of Otolaryngology Head and Neck Surgery, Far Eastern Memorial Hospital, New Taipei City, Taiwan

**Correspondence**
Roger W. Chan, 1999 Guankou Middle Road, Jimei District, Xiamen, Fujian 361023, China
Email: rogerwchan@126.com

## Abstract

**Objectives:** This study developed the Mandarin Chinese version of the Aging Voice Index (AVI), with preliminary validation of the scale for potential clinical applications.

**Study Design:** Scale development.

**Methods:** The experimental procedure involved: (1) cross-cultural adaptation of the original AVI into the Mandarin Chinese version (CAVI); (2) evaluation by expert panel; (3) back translation; (4) pilot testing; (5) development of the final CAVI; (6) scale validation with 68 older adults of 60–89 years old (29 females and 39 males), 34 with voice disorders and 34 age-matched with normal voice. Internal consistency reliability, test–retest reliability, content validity, criterion-related validity, and discriminatory ability (diagnostic accuracy) of the CAVI were evaluated.

**Results:** There were high internal consistency (Cronbach's alpha = 0.9733), high test–retest reliability (intraclass correlation coefficient = 0.9578, $p < 0.01$), high content validity (content validity index = 0.9710), high criterion-related validity (Pearson's $r = 0.9439$, $p < 0.01$ between CAVI and Voice Handicap Index-10; $r = 0.8070$, $p < 0.01$ between CAVI and voice-related quality of life [V-RQOL]), and significant difference in CAVI scores between the two groups with huge effect size ($t(34.69) = -11.59$, Cohen's $d = 2.81$, $p < 0.001$). Receiver operating characteristic analysis revealed a high diagnostic accuracy of the CAVI, with an area under the curve of 0.9974 ($p < 0.001$) and a cut-off score of 12.0 with 100% sensitivity and 97.1% specificity.

**Conclusion:** Our findings suggested that the CAVI could be a reliable and valid standardized self-assessment questionnaire tool for clinical evaluation of the impact of voice problems specifically for Mandarin-speaking older adults. Further studies should explore a full-scale validation of the CAVI for being a standard clinical tool, including for older adults in Mainland China.

**Level of evidence:** 3b (case–control study).

KEYWORDS

cross-cultural adaptation, presbyphonia, quality of life, self-assessment

## 1 | INTRODUCTION

Aging has a significant impact on the structures and functions of various systems of the body, including voice production.[1–4] The prevalence of voice disorders in older adults can be much higher than that in younger adults.[5–8] In addition to objective clinical assessment of vocal function, an individual's own perception of voice problems and their impact on daily life is equally important for a comprehensive management of voice disorders.[9–11] Patient-centered outcome measures and self-assessment rating scales have become increasingly important for objectively evaluating the physical, functional, and psychosocial impact of vocal impairment in individuals with or at-risk of voice disorders.[12–14] Various standardized questionnaires and scales have been developed, including voice handicap index (VHI),[12] voice-related quality of life (V-RQOL),[13] and the shortened VHI-10,[14] with VHI-10 and V-RQOL perhaps being the most widely used clinical tools. Other tailored assessment tools for measuring vocal impairment in specific populations have also been developed, including scales for children,[15,16] and scales for singers and occupational voice users.[17–19]

For older adults, it has been argued that existing voice-related self-assessment tools might not be adequate for evaluating the impact uniquely for them, because only younger participants were targeted for validating existing scales.[11,20,21] To address this issue, the Aging Voice Index (AVI) was developed specifically for evaluating the unique aspects of vocal impairment in older adults, validating the scale on 92 older adults of 60–92 years old (20 with no history of voice problems and 72 with voice disorders) while demonstrating high reliability and validity.[21] There are 23 items (questions) in the AVI, with participants rating each item on a five-point Likert scale (0–4) based on one's self-perception of the frequency of occurrence, yielding a total score of 0–92 where higher scores would indicate more significant impact of vocal impairment.[21]

To date, three cross-cultural adaptations of the AVI have been developed for non-English-speaking populations, including Korean, Persian, and Arabic versions.[22–24] Yet no cross-cultural adaptations of the AVI have been developed for Mandarin speakers so far. Existing Mandarin versions of VHI-10 and V-RQOL developed for clinical applications in Taiwan were not designed nor validated specifically for older adults, either.[25,26] The goal of this study was to culturally adapt the AVI into Mandarin Chinese (CAVI) for potential clinical use, especially for Mandarin-speaking populations in Taiwan, and to conduct a preliminary validation. The hypothesis was that CAVI would be reliable and valid, capable of differentiating older adults with voice disorders from those without.

## 2 | MATERIALS AND METHODS

### 2.1 | Development of Mandarin Chinese version of the AVI

With permission for cross-cultural adaptation from the author of the original AVI, the English AVI was translated into Mandarin Chinese by a bilingual speech-language pathologist (SLP) with 27 years of experience in the management of voice disorders, yielding the first CAVI draft. The principle of translation and adaptation was to preserve the original meaning as much as possible while adapting the wordings specifically for the cultural context of Taiwan.[27,28] This first draft was examined by an expert panel of four bilingual SLPs and two bilingual laryngologists (with 13–40 years of experience in the management of voice disorders) for content validity. The panel evaluated the accuracy of cross-cultural adaptation, including consistency, readability, and cultural appropriateness of the terms and sentences for each of the 23 items, scoring them on a four-point Likert scale with rating scores of 1–4, allowing for computation of the content validity index (CVI).[29,30] By integrating comments and suggestions of the expert panel, the first consensus version of the CAVI was developed with slight modifications in wordings for 6 of the 23 items (for the other 17 items, all panel members agreed that the first draft was perfect). This consensus version was back translated into English by a bilingual translator and was compared with the original English AVI. The back translation showed a strong resemblance to the English version as quantitative single-word analysis showed that 94% of the wordings were identical or equivalent, suggesting that the original meaning was preserved. The CAVI was then subjected to pilot testing.

Pilot testing was conducted to ensure that the CAVI can be comprehended by older adults with no background in voice disorders, to provide feedback on wordings of the scale from a lay perspective,[28] and to estimate the time required for completion. It was completed by five older adults above 60 (three males and two females) who did not participate in the full study. All participants completed the questionnaire smoothly, with one of five participants a bit confused while completing the final five reversely coded items (items 19–23), requiring brief explanation and clarification. The time required to complete the CAVI was 8–10 min, with 1–2 min for instructions and 7–8 min for actual completion of the scale. Based on participant feedback, no changes were necessary for all 23 items, yielding the final version of the CAVI (Table A1).

### 2.2 | Participants

Written informed consents were obtained from all participants. The experimental protocol of this study was approved by the Research Ethics Review Committee of Far Eastern Memorial Hospital in Taiwan (approval number 110038-E). There were 68 participants, with 34 in the disorders group (12 females, 22 males) with a variety of voice disorders (including benign lesions, chronic laryngitis, vocal fold atrophy, vocal fold paralysis and paresis, scarring and sulcus vocalis), and 34 in the control group (17 females, 17 males) age-matched to those in the disorders group (Table 1). There were no significant differences in age between the two groups as a whole, and separately for female and for male participants (Table 1).

Participants in the voice disorders group were recruited from among regular voice clinic patients based on these inclusion criteria: (1) at or above 60 years old according to the definition of

| Participants | Control group age in years (means ± SD) | Disorders group age in years (means ± SD) | Student's *t* (degrees of freedom) | *p* |
|---|---|---|---|---|
| Male | 66.88 ± 6.51 | 67.55 ± 6.32 | −0.3209 (37) | 0.7501 |
| (*n* = 39) | (*n* = 17) | (*n* = 22) | | |
| Female | 69.29 ± 7.43 | 66.67 ± 6.49 | 0.9869 (27) | 0.3324 |
| (*n* = 29) | (*n* = 17) | (*n* = 12) | | |
| Total | 68.09 ± 6.99 | 67.24 ± 6.29 | 0.5290 (66) | 0.5986 |
| (*n* = 68) | (*n* = 34) | (*n* = 34) | | |

*Note*: SD = standard deviations.

**TABLE 1** Age and gender distributions of two groups of participants, with results of Student's *t* tests on difference in age between the groups.

"older people" by the World Health Organization;[31] (2) formal diagnosis of dysphonia by a board-certified laryngologist; (3) self-perceived voice problems, as determined by VHI-10 score (Mandarin version) ≥5;[25,32] (4) being literate to comprehend and complete the questionnaire, as assessed by a short screening interview. Participants in the control group were recruited from local community colleges, churches, and senior citizen centers, based on these inclusion criteria: (1) at or above 60 years old;[31] (2) no history of voice disorders, head and neck diseases, and other medical conditions that may affect voice production; (3) normal voice quality based on informal auditory-perceptual evaluation with the GRBAS rating scale during a short screening interview; (4) no self-perceived voice problems, as determined by a VHI-10 score (Mandarin version) below 5;[25,32] (5) being literate to comprehend and complete the questionnaire.

All participants completed (1) Mandarin version of VHI-10;[25] (2) Mandarin version of V-RQOL;[26] and (3) the CAVI independently. All participants completed the three scales in person on paper, instead of online to avoid any potential effects due to variations in the format of questionnaire administration. The time it took for all participants to complete the CAVI was 10 min on average, with 1–2 min for instructions and 7–8 min for actual completion.

## 2.3 | Data analysis

Statistical analysis was conducted with IBM SPSS version 22.0. A *p* value of <0.05 (two-tailed) was considered statistically significant. Scores obtained for the three scales for all 68 participants were described with means and standard deviations (SD).

## 2.4 | Internal consistency reliability

The Cronbach's alpha coefficient examines whether the items of a scale are measuring the same phenomenon by quantifying the extent to which the items are correlated with one another,[33–35] and was used to assess internal consistency. To evaluate the impact of individual items on internal consistency, item analysis for all items of the CAVI was conducted by (1) corrected item-total correlation analysis (with Pearson's *r*), that is, correlations between each item and the overall CAVI without the item itself (corrected point-biserial correlations); and by (2) successively removing each item at a time while examining changes in Cronbach's alpha for the rest of the CAVI.

## 2.5 | Test–retest reliability

For evaluating test–retest reliability, 21 participants were randomly selected from among all participants to complete the CAVI for a second time after 1 week,[36] consistent with the validation study for the original AVI.[21] To evaluate test–retest reliability, intraclass correlation coefficients (ICC) and their 95% confidence intervals (CIs) were calculated with a consistency, absolute-agreement, single-measures, two-way mixed-effects model.[37,38]

## 2.6 | Content validity

Content validity was quantified by the CVI calculated from rating scores of the expert panel, specifically item-level CVI for each of the 23 items, determined as the number of rating scores of 3 or 4 for that item divided by the total number of experts (6) as an indication of the degree of consistency in opinion among the experts.[29,30] A total CVI was obtained as the average of the item-level CVI values; with total CVI at or above 0.80 indicating acceptable consistency among the experts, or high content validity.[29,30]

## 2.7 | Criterion-related validity

Criterion-related validity of the CAVI was established based on the extent to which CAVI scores were correlated with scores of two existing Mandarin versions of standardized self-assessment tools, VHI-10, and V-RQOL.[25,26] Pearson's correlation coefficient *r* was calculated to determine the correlations between CAVI scores and VHI-10 and V-RQOL scores.

## 2.8 | Independent samples *t* tests

To examine construct validity, independent samples *t* tests were conducted to determine if there were any group differences in CAVI scores, that is, whether there were any significant differences in

CAVI scores between the voice disorders group and the control group.

### 2.9 | Receiver operating characteristic (ROC) analysis

ROC analysis was performed to determine the discriminatory ability or diagnostic accuracy of the CAVI to differentiate older adults with voice disorders from those without, with calculation of the area under the curve (AUC) in the ROC plot.[39–41] Youden's index ($J$ = sensitivity + specificity − 1) analysis was used to determine the optimal cut-off (threshold) score for distinguishing between older adults with and without voice disorders, with the cut-off point established at maximum Youden's index, that is, with a maximum vertical distance between the ROC curve and the reference line in the ROC plot, a widely used approach.[41]

## 3 | RESULTS

### 3.1 | Internal consistency reliability

A Cronbach's alpha value of around or above 0.9 can be considered desirable for clinical assessment tools.[34,35] The Cronbach's alpha value for the CAVI was 0.9733, indicating high internal consistency. Results of corrected item-total correlation analysis showed correlation coefficients from 0.448 to 0.913 for individual items, indicating significant associations between each item and the overall CAVI excluding the item, that items were consistent in measuring the same construct. Results of item analysis showed minimal decreases or increases in Cronbach's alpha values (with magnitude <0.002) with the successive removal of each item, indicating high internal consistency among all 23 items (Table 2).

**TABLE 2** Internal consistency reliability of the CAVI with results of corrected item-total correlation and item analysis (Cronbach's alpha for removal of each item).

| Item | Corrected item-total correlation coefficient[a] | Cronbach's alpha if item was removed from the CAVI[b] |
|---|---|---|
| 1 | 0.799 | 0.972 |
| 2 | 0.817 | 0.972 |
| 3 | 0.864 | 0.971 |
| 4 | 0.886 | 0.971 |
| 5 | 0.866 | 0.971 |
| 6 | 0.767 | 0.972 |
| 7 | 0.837 | 0.972 |
| 8 | 0.913 | 0.971 |
| 9 | 0.907 | 0.971 |
| 10 | 0.827 | 0.972 |
| 11 | 0.641 | 0.973 |
| 12 | 0.895 | 0.971 |
| 13 | 0.770 | 0.972 |
| 14 | 0.697 | 0.973 |
| 15 | 0.827 | 0.972 |
| 16 | 0.877 | 0.971 |
| 17 | 0.520 | 0.974 |
| 18 | 0.870 | 0.971 |
| 19 | 0.618 | 0.973 |
| 20 | 0.821 | 0.972 |
| 21 | 0.695 | 0.973 |
| 22 | 0.448 | 0.974 |
| 23 | 0.677 | 0.973 |

[a]The corrected item-total correlation for each item represented the association between the item and the overall CAVI excluding the item.
[b]Cronbach's alpha of the CAVI with all 23 items = 0.9733.

### 3.2 | Test–retest reliability

ICC analysis was used to investigate the correlation between test scores and retest scores, completed by 21 participants with a 1-week interval in between (5 males and 7 females in the control group; 4 males and 5 females in the disorders group). The ICC value with 95% CIs for the test–retest CAVI scores was 0.9578 [0.8991, 0.9827] ($p$ < 0.001), indicating excellent test–retest reliability among the two completions.[37,38]

### 3.3 | Content validity

Content validity was quantified by item-level CVI values for each of the 23 items. Results showed item-level CVI values of 0.8333 for 4 of the 23 items (items 2, 15, 20, and 22), and 1.000 for the other 19 items of the CAVI, yielding a total CVI value of 0.9710 (Table 3). This finding suggested very high content validity for all 23 items of the CAVI, as CVI values of ≥0.80 can be considered as high content validity, that is, high relevance of the items and high consistency among the experts.[29,30]

### 3.4 | Criterion-related validity

Criterion-related validity of the AVI was established by correlating CAVI scores with scores of Mandarin versions of V-RQOL and VHI-10. Table 4 shows the correlation results for all 68 participants, with Pearson's $r$ = 0.8070 between CAVI and V-RQOL ($p$ < 0.01), 0.9439 between CAVI and VHI-10 ($p$ < 0.01), and 0.8549 between V-RQOL and VHI-10 ($p$ < 0.01). For 34 participants in the disorders group, the correlation results were Pearson's $r$ = 0.6484 between CAVI and V-RQOL ($p$ < 0.01), 0.8668 between CAVI and VHI-10 ($p$ < 0.01), and 0.7453 between V-RQOL and VHI-10 ($p$ < 0.01). These Pearson's $r$ values can be considered as high to very high levels of correlation.[42]

**TABLE 3** Item-level content validity index (CVI) values for 23 items of the CAVI.

| Items | Questions (back translated into English) | CVI |
|---|---|---|
| 1 | People ask, "What is wrong with your voice?" | 1.000 |
| 2 | Because of my voice problem, others often cannot hear me. | 0.8333 |
| 3 | My voice problem frustrates me. | 1.000 |
| 4 | My voice problem leads me to run out of air when I talk. | 1.000 |
| 5 | I feel frustrated with changes in my voice. | 1.000 |
| 6 | Other people thought I am sick because of my voice problem. | 1.000 |
| 7 | Because of my voice problem, I talk less. | 1.000 |
| 8 | I feel effortful when I speak. | 1.000 |
| 9 | I am annoyed when my voice is not working. | 1.000 |
| 10 | My voice problem affects what I want to do. | 1.000 |
| 11 | Other people judge me negatively because of my voice. | 1.000 |
| 12 | I am worried about my voice. | 1.000 |
| 13 | Because of my voice problem, I have to stop participating in activities that are important to me (e.g., singing, volunteering, working, etc.) | 1.000 |
| 14 | Because of my voice problem, other people will finish my talk for me. | 1.000 |
| 15 | I do not like the way my voice sounds. | 0.8333 |
| 16 | My voice problem makes me feel sad. | 1.000 |
| 17 | I completely lost my voice. | 1.000 |
| 18 | I feel obstructed due to my voice problem. | 1.000 |
| 19 | I can talk on the phone as much as I want. | 1.000 |
| 20 | I like the way my voice sounds. | 0.8333 |
| 21 | I can talk as much as I want, for as long as I want. | 1.000 |
| 22 | My family and friends understand what I talk about. | 0.8333 |
| 23 | My voice is as good as I want it to be. | 1.000 |
| Total CVI for the CAVI | | 0.9710 |

### 3.5 | Independent samples *t* tests

Mean CAVI scores were 2.35 ± 2.99 for the control group and 39.97 ± 18.69 for the voice disorders group, with a significant difference between the two groups with a huge effect size ($t(34.6921) = -11.5876$, Cohen's $d = 2.8104$, $p < 0.001$),[43] supporting the construct validity of the CAVI.

### 3.6 | ROC analysis

Figure 1 shows ROC curves based on CAVI scores and V-RQOL scores of all 68 participants. For CAVI scores, the AUC was 0.9974 (95% CIs: 0.9910–1.0000) ($p < 0.001$), indicating an outstanding discriminatory ability or diagnostic accuracy.[40,41] Youden's index analysis identified an optimal CAVI cut-off score of 12.0, with 100% sensitivity and 97.1% specificity.[39–41] For V-RQOL scores, the AUC was 0.9650 (95% CIs: 0.9218–1.000) ($p < 0.001$), indicating a lower but still outstanding discriminatory ability or diagnostic accuracy.[40,41] Youden's index analysis identified an optimal V-RQOL cut-off scores of 10.5, with 97.1% sensitivity and 85.3% specificity.[39–41] As V-RQOL scores are expressed in integers, the V-RQOL cut-off score was rounded off as 11.0.

**TABLE 4** Correlations among CAVI scores, V-RQOL scores, and VHI-10 scores for all participants ($n = 68$).

| Pearson's *r* | CAVI | V-RQOL | VHI-10 |
|---|---|---|---|
| CAVI | – | | |
| V-RQOL | 0.8070** | – | |
| VHI-10 | 0.9439** | 0.8549** | – |

**$p < 0.01$ (two-tailed).

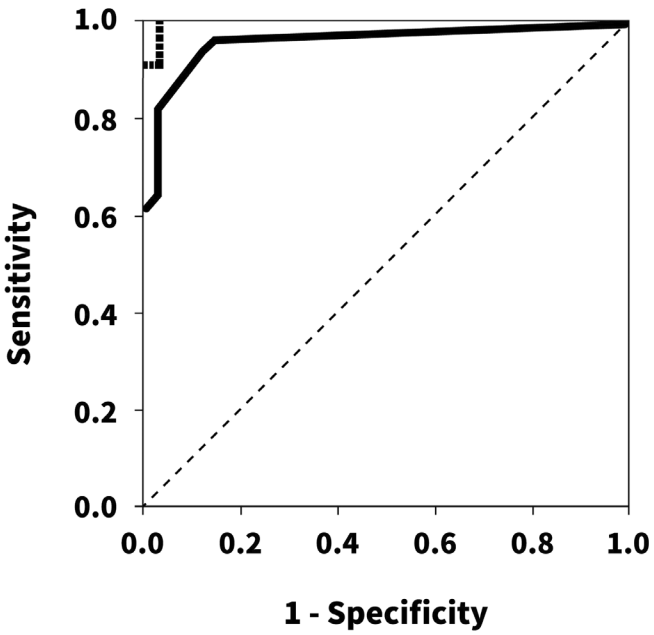


**FIGURE 1** Receiver operating characteristic (ROC) curves for Mandarin Chinese version of Aging Voice Index (CAVI) scores (dotted line) and for Mandarin version of Voice-Related Quality of Life (V-RQOL)[26] scores (solid line) to distinguish between the voice disorders group ($n = 34$) and the control group ($n = 34$) (diagonal dashed line is the reference line with area under the curve = 0.5).

## 4 | DISCUSSION

### 4.1 | Reliability of the CAVI

Cronbach's alpha values of 0.9 or above can be considered desirable for clinical assessment tools.[34,35] Cronbach's alpha values of 0.949, 0.832, 0.83, 0.954, and 0.9733 were found for the original English AVI,[21] KAVI,[22] P-AVI,[23] A-AVI,[24] and CAVI in this study, respectively.

**TABLE 5** Summary of scores (means ± variability) for different versions of the Aging Voice Index across participant groups.

| Versions of AVI | AVI scores (means ± variability[a]) | | |
|---|---|---|---|
| | Control group (sample size) | Voice disorders group (sample size) | Statistical analysis results |
| Original AVI[21] | 6.050 ± 3.167 | 35.652 ± 3.40[b] | Least mean squares with Tukey test (results not specified) |
| | ($n = 20$) | ($n = 72$) | $p < 0.0001$ |
| KAVI[22] | 8.51 ± 1.685 | 33.68 ± 9.561 | $t = -23.004$ (effect size not specified) |
| | ($n = 100$) | ($n = 111$) | $p < 0.001$ |
| P-AVI[23] | 1.05 ± 1.05 | 38.70 ± 10.71 | Mann–Whitney $U$ test ($U$ not specified) |
| | ($n = 20$) | ($n = 85$) | $p < 0.001$ |
| A-AVI[24] | 2.84 ± 5.441 | 40.68 ± 22.195 | (results not specified) |
| | ($n = 77$) | ($n = 82$) | $p < 0.001$ |
| CAVI | 2.3529 ± 2.9938 | 39.9706 ± 18.6912 | $t$ (34.6921) = −11.5876 |
| (this study) | ($n = 34$) | ($n = 34$) | Cohen's $d = 2.8104$ $p < 0.001$ |

Abbreviations: AVI, Aging Voice Index; KAVI, Korean version of AVI; P-AVI, Persian version of AVI; A-AVI, Arabic version of AVI.
[a]Variability was expressed in standard deviations for all studies, except for the original AVI where variability was expressed in standard errors.
[b]Only mean AVI scores for individual diagnostic categories were reported in Etter et al.,[21] but not mean scores for all 72 participants with voice disorders. The mean AVI scores ± standard errors here were calculated from their reported figures.

These values of Cronbach's alpha indicated high internal consistency, especially for the original version, Arabic version, and the CAVI. Corrected item-total correlations (corrected point-biserial correlations) indicated that all individual items of the CAVI were well correlated with the entire scale without the items themselves (Table 2). Item analysis results also suggested high internal consistency among all 23 items of the CAVI, with only minimal changes in Cronbach's alpha upon successively removing each item (Table 2).

For evaluating test–retest reliability, the time interval separating two completions of standardized questionnaire tools is typically 1 to 2 weeks.[36] An ICC value (with 95% CIs) of 0.9578 (0.8991, 0.9827) ($p < 0.001$) was found between the first and the second completions separated by 1 week, suggesting that the two completions of the CAVI by the same 21 participants were highly reliable with each other, indicating excellent test–retest reliability. This was comparable to those for other versions of AVI (with ICC values of 0.952, 0.851, 0.95, 0.987 for the original AVI, KAVI, P-AVI, and A-AVI, respectively).

## 4.2 | Validity of the CAVI

For evaluating content validity of standardized questionnaire tools, a minimum of five experts in a panel has been recommended.[29] There were six experts in this study, yielding item-level CVI of 1.0 for 19 of the 23 items, which indicated perfect agreement among the experts for those items; and 0.8333 for the other 4 items, still considered as high content validity.[29,30] A total CVI value of 0.9710 was found for the CAVI, suggesting that all items were considered highly relevant and valid among members of the expert panel (Table 3). Regrettably, CVI values were not reported for all other versions of AVI to allow for quantitative comparisons with our current findings.

On criterion-related validity, high to very high levels of correlation were observed between CAVI scores and the scores of two currently available standardized Mandarin-version questionnaires, V-RQOL and VHI-10 (Pearson's $r$ ranging from 0.8070 to 0.9439 for all 68 participants; 0.6484 to 0.8668 for 34 participants with voice disorders) (Table 4), indicating high criterion-related validity for the CAVI.[42] These levels of correlation were comparable to those of other versions of AVI being correlated to their respective versions of V-RQOL or VHI-10, including the original AVI ($r = 0.879$ with V-RQOL), KAVI ($r = 0.994$ with V-RQOL), P-AVI ($r = 0.86$ with V-RQOL), and A-AVI ($r = 0.89$ with VHI-10).

On construct validity, results of independent samples $t$ tests showed that the disorders group demonstrated significant higher CAVI scores than the control group (mean scores of 39.97 vs. 2.35), with the difference being highly significant with a huge effect size (Cohen's $d = 2.8104$).[43] This was consistent with previous versions of AVI, where significant differences in AVI scores were also found between older adults with and without voice disorders. The range of scores was also consistent with previous versions of AVI, with mean scores at or below 8.51 for older adults with normal voice in other versions of AVI, and mean scores ranging from 24.50 to 40.68 for those with voice disorders in the original AVI, KAVI, P-AVI, and A-AVI (Table 5).

## 4.3 | ROC analysis

Results of ROC analysis examining the discriminatory ability or diagnostic accuracy of the CAVI showed an AUC of 0.9974, indicating outstanding diagnostic accuracy for distinguishing between the two participant groups.[40] AUC values can range from 0 to 1.0, with

**TABLE 6** Distributions of CAVI scores in the two participant groups.

| Range of AVI scores | 0–10 | 10–20 | 20–30 | 30–40 | 40–50 | 50–60 | 60–70 | 70–80 | >80 |
|---|---|---|---|---|---|---|---|---|---|
| Control group ($n = 34$) | 33 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Voice disorders group ($n = 34$) | 0 | 6 | 4 | 7 | 8 | 3 | 3 | 2 | 1 |

0 indicating a perfectly inaccurate test, 1.0 indicating a perfectly accurate test, 0.5 suggesting no discrimination, 0.7 to 0.8 considered acceptable, 0.8 to 0.9 considered excellent, and >0.9 considered as outstanding discriminatory ability.[40,41] An ROC plot depicts sensitivity in the *y*-axis (true positives) and (1 − specificity) in the *x*-axis (false positives), with an ROC curve at the upper left corner indicating perfect classification.[39–41] Figure 1 shows a higher discriminatory ability for the ROC curve based on CAVI scores than that based on V-RQOL scores of the 68 participants, corresponding to the difference in their AUCs (0.9974 vs. 0.9650). This suggested that the CAVI might be a more sensitive self-assessment tool than the V-RQOL for older adults, consistent with the argument that the CAVI could be more tailored than other existing assessment tools for evaluating the impact of vocal impairment on quality of life in older adults.[11,20,21]

Youden's index analysis revealed a preliminary optimal cut-off score of 12.0 for the CAVI for distinguishing between older adults with and without voice disorders. As the total scores of CAVI can range from 0 to 92, a cut-off score of 12.0 may seem to be quite low within the score range, and could be related to the distributions of scores for our participants. Table 6 shows the score distributions, where it can be seen that the scores of almost all (33 of 34) participants in the control group were in the range of 0–10; whereas for the voice disorders group, the scores of a majority of participants (24 of 34) were in the range of 30 to 83. Given such vastly different distributions of scores for the two participant groups, it was not surprising to yield a low cut-off score of 12.0 for the CAVI. Regrettably, no ROC analysis was conducted in the validation studies for the other versions of AVI.[21–24]

An obvious limitation of the current study was the relatively small sample size (68 participants), which was not adequate for a full validation of the scale. This study should therefore be considered as preliminary validation, with a large-scale study needed for fully validating the CAVI. A comprehensive review article on psychometric validation studies of patient-reported outcomes measures (standardized questionnaires and scales) suggested that "subjects to items ratio" has been a frequently used index for determining the required sample size for validation studies in the literature,[44] with the recommended ratio ranging from 2 to 20 subjects per item.[45,46] With 68 participants and 23 CAVI items in this study, our subjects to items ratio was 2.9565, near the lower limit of the recommended range and likely resulting in insufficient statistical power for minimizing type II error. A second limitation was that participants in the disorders group had a variety of voice disorders, where differences in their impacts could be a confounding factor affecting the CAVI scores. A third limitation was that the inclusion criteria for participants in the control group were primarily based on the self-perception of normal voice (as determined by VHI-10 scores) and by auditory-perceptual evaluation of normal voice quality, without more objective voice assessment. Future studies validating the CAVI should involve objective assessment approaches such as endoscopic examination and acoustic analysis as part of the participant recruitment process.

## 5 | CONCLUSION

Preliminary findings suggested that the CAVI could be a proper self-assessment tool for evaluating the specific and unique impact of vocal impairment in Mandarin-speaking older adults, capable of distinguishing between older adults with and without voice disorders, with a preliminary optimal cut-off score of 12.0. Further studies with a larger sample size are recommended for fully validating the CAVI. In addition, further studies should also target adaptation and validation of the CAVI specifically for Mandarin-speaking older adults in China, as there are certain linguistic and cultural differences in the usage of Mandarin in Taiwan versus in Mainland China.

### ACKNOWLEDGMENTS

We would like to thank Sherry Fu, Chi-Fen Chang, Chin-Wen Chang, and Tzu-Yu Hsiao for serving in the expert panel evaluating the CAVI draft.

### FUNDING INFORMATION

This work was supported by a grant from Xiamen Medical College (Science and Technology Project Fund no. KPT2020-02).

### ORCID

*Roger W. Chan* https://orcid.org/0000-0002-7135-9702
*Chi-Te Wang* https://orcid.org/0000-0002-9999-1291

### REFERENCES

1. Sataloff RT, Rosen DC, Hawkshaw M, Spiegel JR. The aging adult voice. *J Voice*. 1997;11(2):156-160.
2. Sharma G, Goodwin J. Effect of aging on respiratory system physiology and immunology. *Clin Interv Aging*. 2006;1(3):253-260.
3. Kendall K. Presbyphonia: a review. *Curr Opin Otolaryngol Head Neck Surg*. 2007;15(3):137-140.
4. Gregory ND, Chandran S, Lurie D, Sataloff RT. Voice disorders in the elderly. *J Voice*. 2012;26(2):254-258.
5. Golub JS, Chen PH, Otto KJ, Hapner E, Johns MM 3rd. Prevalence of perceived dysphonia in a geriatric population. *J Am Geriatr Soc*. 2006;54(11):1736-1739.
6. Cohen SM, Kim J, Roy N, Asche C, Courey M. Direct health care costs of laryngeal diseases and disorders. *Laryngoscope*. 2012;122(7):1582-1588.

7. Roy N, Stemple J, Merrill RM, Thomas L. Epidemiology of voice disorders in the elderly: preliminary findings. *Laryngoscope*. 2007;117(4):628-633.
8. Roy N, Kim J, Courey M, Cohen SM. Voice disorders in the elderly: a national database study. *Laryngoscope*. 2016;126(2):421-428.
9. Murry T, Rosen CA. Outcome measurements and quality of life in voice disorders. *Otolaryngol Clin N Am*. 2000;33(4):905-916.
10. Hogikyan ND, Rosen CA. A review of outcome measurements for voice disorders. *Otolaryngol Head Neck Surg*. 2002;126(5):562-572.
11. Etter NM, Stemple JC, Howell DM. Defining the lived experience of older adults with voice disorders. *J Voice*. 2013;27(1):61-67.
12. Jacobson BH, Johnson A, Grywalski C, et al. The voice handicap index (VHI): development and validation. *Am J Speech Lang Pathol*. 1997;6(3):66-70.
13. Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice*. 1999;13(4):557-569.
14. Rosen CA, Lee AS, Osborne J, Zullo T, Murry T. Development and validation of the voice handicap index-10. *Laryngoscope*. 2004;114(9):1549-1556.
15. Zur KB, Cotton S, Kelchner L, Baker S, Weinrich B, Lee L. Pediatric voice handicap index (pVHI): a new tool for evaluating pediatric dysphonia. *Int J Pediatr Otorhinolaryngol*. 2007;71(1):77-82.
16. Boseley ME, Cunningham MJ, Volk MS, Hartnick CJ. Validation of the pediatric voice-related quality-of-life survey. *Arch Otolaryngol Head Neck Surg*. 2006;132(7):717-720.
17. Cohen SM, Jacobson BH, Garrett CG, et al. Creation and validation of the singing voice handicap index. *Ann Otol Rhinol Laryngol*. 2007;116(6):402-406.
18. Cohen SM, Statham M, Rosen CA, Zullo T. Development and validation of the singing voice handicap-10. *Laryngoscope*. 2009;119(9):1864-1869.
19. Nanjundeswaran C, Jacobson BH, Gartner-Schmidt J, Verdolini AK. Vocal fatigue index (VFI): development and validation. *J Voice*. 2015;29(4):433-440.
20. Pernambuco L, Espelt A, Costa de Lima K. Screening for voice disorders in older adults (RAVI)-part III: cutoff score and clinical consistency. *J Voice*. 2017;31(1):117.e17-e22.
21. Etter NM, Hapner ER, Barkmeier-Kraemer JM, Gartner-Schmidt JL, Dressler EV, Stemple JC. Aging voice index (AVI): reliability and validity of a voice quality of life scale for older adults. *J Voice*. 2019;33(5):807.e7-807.e12.
22. Bae IH, Sung ES, Lee JC. Validity and reliability of Korean version of the aging voice index (KAVI). *J Korean Soc Laryngol Phoniatr Logoped*. 2019;30(1):21-27.
23. Ahmadi A, Hosseinifar S, Faham M, et al. Translation, validity, and reliability of the Persian version of the aging voice index. *J Voice*. 2021;35(2):327.e13-e21.
24. Alkhunaizi AA, Almohizea M, Bukhari M, Alhajress RI, Malki KH, Mesallam TA. Validation and cultural adaptation of the Arabic version of the aging voice index. *J Voice*. 2022;S0892-1997(22):00032-7.
25. Wang NM, Huang KY, Su MC, Hsin CH. Development and validation of the voice handicap index-10 in Mandarin Chinese. *J Taiwan Otolaryngol Head Neck Surg*. 2011;46(4):190-196.
26. Huang YA, Hsu TA, Fu S, Wang LM. Reliability and validity of the mandarin version of the voice-related quality of life (MV-RQOL) measure. *J Voice*. 2021;S0892-1997(21):00345-3.
27. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417-1432.
28. Hilton A, Skrutkowski M. Translating instruments into other languages: development and testing processes. *Cancer Nurs*. 2002;25(1):1-7.
29. Lynn MR. Determination and quantification of content validity. *Nurs Res*. 1986;35(6):382-385.
30. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30(4):459-467.
31. World Health Organization (WHO). UN Decade of Healthy Ageing: Plan of action 2021–2030. Accessed February 24, 2021. https://cdn.who.int/media/docs/default-source/decade-of-healthy-ageing/decade-proposal-final-apr2020-en.pdf
32. Wu CH, Chan RW. Effects of a 6-week straw phonation in water exercise program on the aging voice. *J Speech Lang Hear Res*. 2020;63(4):1018-1032.
33. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297-334.
34. Bland JM, Altman DG. Cronbach's alpha. *BMJ*. 1997;314:572.
35. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53-55.
36. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: a Practical Guide to their Development and Use*. 5th ed. Oxford University Press; 2015.
37. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
39. Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett*. 2006;27(8):861-874.
40. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315-1316.
41. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr*. 2007;96(5):644-647.
42. Hinkle DE, Wiersma W, Jurs SG. *Applied Statistics for the Behavioral Sciences*. Vol 663. 5th ed. Houghton Mifflin; 2003.
43. Sawilowsky SS. New effect size rules of thumb. *J Mod Appl Stat Methods*. 2009;8:597-599.
44. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes*. 2014;12:176-185.
45. Hair JF, Anderson RE, Tatham RL, Black WC. *Multivariate Data Analysis: with Readings*. 4th ed. Prentice-Hall; 1995:757.
46. Kline P. *Psychometrics and Psychology*. Academic Press; 1979:381.

**How to cite this article:** Liao C-E, Chan RW, Zhuang P, et al. Mandarin Chinese version of the Aging Voice Index. *Laryngoscope Investigative Otolaryngology*. 2023;8(4):980-988. doi:10.1002/lio2.1117

## APPENDIX A

**TABLE A1** Mandarin Chinese version of the AVI (CAVI) in traditional Chinese characters

| **老年嗓音指數量表** | **0** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|---|
| 1. 別人會問:「你的聲音怎麼了?」 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 2. 因為我的嗓音問題，別人常常聽不到我的聲音 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 3. 我的嗓音問題讓我覺得沮喪 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 4. 我的嗓音問題讓我說話時覺得氣不夠 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 5. 嗓音的改變讓我覺得沮喪 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 6. 別人會因為我的嗓音問題，以為我生病了 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 7. 因為我的嗓音問題，我比較少說話 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 8. 我說話時覺得費力 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 9. 嗓音狀況不好時讓我覺得困擾 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 10. 我的嗓音問題影響到我想做的事情 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 11. 別人會因為我的嗓音而對我有負面的評價 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 12. 我擔心我的嗓音狀況 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 13. 因為我的嗓音問題，我必須停止一些對我來說是重要的活動 (例如:唱歌、擔任志工、工作...等等) | 從不 | 很少 | 有時 | 經常 | 總是 |
| 14. 因為我的嗓音問題，別人會幫我把話說完 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 15. 我不喜歡我的聲音聽起來的樣子 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 16. 我的嗓音問題讓我覺得難過 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 17. 我完全失聲了 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 18. 我的嗓音問題讓我覺得受到阻礙 | 從不 | 很少 | 有時 | 經常 | 總是 |
| **雖然我有嗓音上的問題...** | **4** | **3** | **2** | **1** | **0** |
| 19. 我可以想講多少電話就講多少 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 20. 我喜歡我的聲音聽起來的樣子 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 21. 我可以想講多少話就講多少，想講多久就講多久 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 22. 我的家人和好朋友聽得懂我說什麼 | 從不 | 很少 | 有時 | 經常 | 總是 |
| 23. 我的嗓音和我希望的一樣好 | 從不 | 很少 | 有時 | 經常 | 總是 |