



Published in final edited form as:

*Logoped Phoniatr Vocol.* 2009 ; 34(3): 117–127. doi:10.1080/14015430903062712.

## THE INABILITY TO PRODUCE SOFT VOICE (IPSV): A TOOL TO DETECT VOCAL CHANGE IN SCHOOL TEACHERS

Angela E. Halpern<sup>1</sup>, Jennifer L. Spielman<sup>1</sup>, Eric J. Hunter<sup>1</sup>, and Ingo R. Titze<sup>1,2</sup>

<sup>1</sup>National Center for Voice and Speech, Denver, CO USA

<sup>2</sup>Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, IA USA

### Introduction

A healthy voice is essential for a teacher's livelihood. However, due to heavy vocal demands, teachers are at a high risk for developing voice disorders (1–5). Voice problems typically reported by teachers include fatigue, discomfort, increased effort while speaking, hoarseness, breathiness, change in voice quality, and dry throat (2, 3, 6–8). The consequences for teachers of such problems may include lost work days (2, 9, 10), changing lesson plans (8), reduced teaching activities (7), frustration (8), and concern about the future of their career (2, 9).

Many teachers report “fatigue” as one of their main voice problems (4, 8, 11). Clinically, vocal fatigue may be considered to be a worsening or weakening of the voice following prolonged use, resulting in a variety of symptoms (i.e., increased effort for phonation, discomfort, pain, tension, reduced pitch range, and changes in voice quality) which tend to resolve following a period of phonation rest (12–15). In some individuals, the effects of vocal fatigue are clearly perceptible in terms of degraded voice quality. However, clinic patients complaining of vocal fatigue may also present with perceptually normal voices that change during the course of the day, making measurement of the problem challenging.

Performing phonation tasks at both a low intensity and high pitch has been proposed as one simple and sensitive method of detecting vocal fold swelling (24), a potential physiological manifestation of laryngeal tissue fatigue. Distinguished from laryngeal muscle fatigue (26), laryngeal tissue fatigue may occur as a result of excessive tissue vibration at the level of the vocal folds, which has been shown in animal models to result in injury to the lamina propria (27). This excessive tissue vibration may lead to vocal fold disorders (28) and/or vocal fold edema or swelling (29). Edema at the level of the vocal folds may affect the mucosal wave, making soft, high phonation difficult, and thus result in a worse rating of the ability to produce soft, high pitches (30). Soft, high phonation has been shown to identify vocal fold swelling in singers when rated by a third party judge (24), and when used as a self-rating tool (30).

The value of such a simple, efficient, perceptive measure for detecting voice problems is that it can be administered without complex instrumentation by clinicians as well as patients (24). This type of tool could be especially helpful for enabling teachers to identify the effects of excessive voice use (which can result in fatigue) (25), potentially allowing them to engage in brief voice rest or vocal economy strategies to prevent further fatigue and promote voice recovery. In addition, early identification of voice problems could prompt individuals to seek professional help to correct the problem before symptoms become worse or pathology develops (24).

In an initial test of such a measure, Hunter and Titze (25) evaluated 87 full-time teachers for vocal fatigue recovery using a standardized measure of the inability to produce soft voice (IPSV).. The IPSV tasks included the following tasks: [1] sustaining /i/ as softly as possible on a comfortable pitch; [2] gliding from low to high pitch on the vowel /i/ as softly as possible; [3] repeating the syllable /i/ in a high pitched staccato fashion; and [4] singing the first two phrases of the “Happy Birthday” song, soft and high pitched (“Happy Birthday to you, Happy Birthday to you”). Participants were asked to perform these tasks, and then rate themselves on a 1–10 scale (1 = no problems with the tasks, 10 = significant problems). The authors demonstrated that the IPSV was an effective indicator of vocal fatigue recovery, which took place over 12–18 hours following a 2-hour vocal loading task. Therefore, these results also offer the possibility that the IPSV could be used as an indicator of the beginnings of fatigue.

As stated above, teachers or clients complaining of fatigue may present with perceptually normal voices. Thus, the IPSV may be a simple tool that could be completed quickly and easily by teachers. The IPSV may also serve as a “common” measure for the teachers and clinicians, providing the clinician a method to identify progression in their patients toward or away from fatigue, as well as a link between what is happening in the treatment room versus in the “real world”.

To examine the clinical utility of such a measure, it would be valuable to understand the degree of consistency between self and clinician’s ratings of the ability to produce soft voice, something which is currently unclear. As a first step, we chose unstable or poorly targeted vocal fundamental frequency ( $F_0$ ) and/or vocal intensity as principle components influencing these self ratings. This is based primarily on the hypothesis, described above, that difficulty with soft, high phonation may be indicative of vocal fold edema following excessive voice use (30). Additionally, although the relationship between reports of vocal fatigue and changes in acoustic measures of voice have not been clearly established (REFS16–18), vocal intensity and/or  $F_0$  have been associated with excessive voice use (Laukkanen, Jarvinen (17), Lehto (18)).

The current study determined the correlation between clinician and teacher IPSV ratings of soft voice samples. Specifically, the study investigated if [1] clinicians could detect voice changes by analyzing IPSV tasks previously recorded and rated by teachers; and [2] the changes detected by the clinicians would be similar to those detected by the teachers themselves. The two specific voice characteristics which were rated were fundamental frequency and intensity. These two were chosen because they are not only the hypothesized

key components of the IPSV rating, but they are also the two variables of the voice that would be available to both the teachers and the clinicians. Two speech clinicians with specialized training in voice disorders conducted a post hoc rating of the soft voice samples that each of the teachers rated. The clinicians used a 1–10 scale for their IPSV ratings, as did the teachers in the original data set.

## Materials and methods

### Material

The NCVS database on voice dosimetry of teachers was used for analysis. Included in this study were ten teachers (6 females, 4 males) who wore the NCVS vocal dosimeter (32) for 14 days each (although one teacher had data for only 13 days). Dosimetry data were collected via an accelerometer attached to each teacher's neck. The dosimeter can record up to 24 hours of data in a single session (32). A full description of the dosimeter can be found in Popolo et al. (33).

### Teachers

Teachers were recruited from primary and secondary schools in the Denver, Colorado, area after permission was obtained from the school district and principal. All teachers signed consent forms approved by the Colorado Multiple Institutional Review Board. See Table 1 for a full description of teacher characteristics.

**Teacher Baseline Procedures**—Teachers first participated in a baseline evaluation during a period of time when they were not teaching, typically during the summer. During the baseline evaluation, teachers completed an adapted version of the Vocal Health Questionnaire (34) and the Voice Handicap Index (35). They also participated in a videostroboscopic exam and voice and speech tasks. Teachers who passed the baseline were considered appropriate for the study. Teachers were excluded from the study if they could not complete the laryngeal exam, or if the laryngeal examination indicated that the teacher required immediate medical and/or therapeutic attention. No other exclusion criteria were used.

**Teacher Experimental Procedures**—During a period of time when the teachers were actively teaching, the dosimeter was calibrated to each individual's voice (see Svec (36) for a description of the calibration procedure), and teachers were trained in the use of the dosimeter. Within a few days of this training, the teachers were asked to wear the dosimeter from the time they awoke until the time they went to bed for 14 consecutive days. Skin acceleration level (dB SAL), which correlates to dB SPL (37), and fundamental frequency (F0) were recorded continuously in 30 millisecond (ms) time intervals during the entire time the dosimeter was attached and turned on.

While wearing the dosimeter, teachers were asked to perform soft phonation tasks approximately every 2 hours and rate their voices immediately following the four IPSV tasks. The soft phonation tasks were: [1] a sustained /i/; [2] pitch glide (low to high) on the vowel /i/; [3] five repetitions of the syllable /i/ in a staccato like fashion; and [4] the first two

phrases of the “Happy Birthday” song. Teachers were asked to perform the last three tasks at a high pitch. A comfortable pitch was used for the first IPSV task to more easily help the teachers focus on the IPSV and get their voices as soft as possible. To assist with their self-judgments, the teachers were instructed (in lay terms) to listen for instabilities such as roughness/breathiness, periods of aphonia, voice breaks, delayed voice onset, and general difficulties in producing soft voice. The number and severity of the impairments were coded into a single rating on a numeric scale of 1 – 10. Teachers were told that a rating of 1 was “ideal soft-voice production”, and 10 was “a complete inability to produce soft voice”. Although not used in the current study, the teachers were also asked to provide a 1–10 rating (as a part of a separate voice rating task) regarding both the amount of effort and discomfort they felt.

Data from the dosimeter were transferred to a computer. Mean, median, mode, standard deviation, variance, 1st and 3rd quartile, skewness, kurtosis and distribution information were calculated in Matlab on dB SAL and F0 for each of the four IPSV tasks produced. See Figure 1 for a display of one hour of dosimetry data for running speech.

**Clinician Experimental Procedures:** Two speech clinicians rated each of the IPSV tasks for each teacher, blinded to the teacher’s own ratings. The two clinicians coordinated their ratings, generating one score for each IPSV task. The dosimeter by design does not record the full radiated acoustic signal, but rather a reduced signal by calculating F0 and dB SAL produced by the teacher every 30 ms. Thus the clinicians rated the graphical representations of F0 and dB SAL (Figure 2), accompanied by synthesized acoustic representations of the tasks, in which F0 and intensity were reproduced auditorily according to Figure 2. These audio signals and visual displays also provided cursory information regarding smoothness and stability of the voice (e.g. steadiness, phonation and pitch breaks). The synthesized representations were created by using a digital voltage controlled oscillator, which could produce varied F0 and dB SAL. The synthesized version of the F0 and dB SAL data supplemented the visual displays of the F0 and dB SAL.

The steps involved in rating each IPSV task were as follows:

1. Using a customized script, each IPSV file was edited to eliminate extraneous talking and noise, and to delineate only the four IPSV tasks (see Figure 2).
2. A “gold standard” for each subject was created by the clinicians, to act as an anchor for the clinician ratings. This “gold standard” was created by selecting the best production of each single task from the entire set of a teacher’s IPSV files. “Best” was defined as the “softest” (task 1) and “softest, highest” production (tasks 2–4), and was determined using both the statistics generated by a second customized script (F0 and dB SAL) as well as the visual and synthesized auditory signal information. The “best productions” for each of the 4 tasks were then pasted together as the “gold standard” for that teacher (Figure 3). Gold standard files were designed to represent the best possible soft phonation that could theoretically be achieved by the individual teacher. It was anticipated that a teacher would rarely, if ever, perform his or her best on all 4 tasks in one single

rating session, and so a score of “0” was applied to the gold standard rather than “1”.

3. Using a third customized rating script, the clinicians then rated each soft phonation file for each teacher. The script allowed the clinicians to view each soft phonation file (Figure 2), listen to the synthesized acoustic signal for that file, and compare that file to the previously selected gold standard for that teacher (Figure 3). There were no restrictions regarding the number of times the clinicians could listen to each acoustic signal. The clinicians assigned a score of 1 – 10 (10 being the poorest in comparison to the gold standard) to each file based on a specific set of criteria. It has been previously demonstrated that when comparing ratings conducted by others to self ratings, it is important to rate the same parameter and to use similar rating scales. Karnell et al. (38) determined that higher correlations between patient and clinician ratings are achieved when both clinician and patient are rating the same parameter (i.e. dysphonia), than when they are rating different parameters (i.e. clinician is rating dysphonia and patient is rating effect of dysphonia on quality of life). Thus, to be consistent with the teachers’ ratings, the clinicians also used a 1–10 scale. A few examples of the criteria are as follows: A rating of a “1” was given if, “everything was at best performance for all 4 tasks, and at the highest/quietest end of the range (relative to each person).” A rating of “6” was given for “significantly lowered pitch relative to intact or possibly increased loudness, obvious difficulty completing one task, more variability in steadiness, and/or an episode of aphonia. Finally, a rating of “10” was given if the teacher produced phonation that was “too loud and/or too low as compared to the gold standard, if the teacher had difficulty with tasks, and if aphonia was present”.

## Analysis

A linear weighted Cohen’s kappa measure as presented in Fleiss et al. (39) was calculated to investigate the agreement between the teacher’s self ratings (IPSVT) and the clinician’s ratings (IPSVC) for each subject. Fleiss (39) describes Cohen’s weighted kappa as “kappa measure of interrater agreement to the case where the relative seriousness of each possible disagreement could be quantified” (p. 608). Means, standard deviations and modes were calculated for each set of teacher self-ratings, as well as each set of clinician ratings for each teacher. *Difference scores* were calculated by subtracting the teacher’s rating (IPSVT) from the clinician’s rating (IPSVC) for each IPSV task. This score determined what percent of IPSVT and IPSVC files were given the same rating (a difference score of 0), and what percent were given a rating that was within one point (a difference score of 1). The mean and standard deviation of these difference scores were then calculated for each teacher. The mode was also calculated because previous literature (40) has demonstrated the utility of including mode in analyses as it reveals the most frequently occurring rating or score and, unlike average, would not be skewed by outliers. Finally, the stability of the IPSV ratings over 2 weeks of ratings was evaluated using measurements of slope, to determine the change over time. Slope change was calculated for the 2 weeks for each teacher, and then the slopes were averaged for all teachers.

## Results

The teachers completed daily self-ratings of soft voice (IPSV) approximately every 2 hours throughout each of the 14 days. This resulted in an average of 90.8 (stdev:16) IPSV ratings per teacher for the 14 days for all but one of the teachers (M056 had 13 days of ratings), with a range of 70–129 ratings. Fewer ratings were due to equipment failure, not wearing the dosimeter consistently, or less time spent awake.

The Cohen's kappa revealed that the individual agreement between the IPSVT and the IPSVC ratings for each teacher was not statistically significant ( $p > 0.05$ ). The group mean and standard deviations of the IPSVT and IPSVC ratings were 4.1 (1.5) and 3.1 (1.1) respectively, with an average absolute difference score of 1.7 (1.4). The IPSVT mean ratings were below 5 for 7 of the 10 teachers, and IPSVC mean ratings were below 5 for all 10 teachers. The IPSVT ratings ranged from 1.7 to 7.8 and the IPSVC ratings ranged from 2.3 to 4.9. IPSVT and IPSVC ratings were within 0 or 1 point of each other more than 65% of the time for 3 teachers; 30 - 65% of the time for 5 teachers; and less than 30% of the time for 2 teachers. The group mode and standard deviations of the IPSVT and IPSVC ratings for all teachers were 3.4 (2.3) and 2.9 (0.9) respectively. See Table 2 for group and individual results. In addition to the results above, mean group and standard deviation slope measurements of 0.07 (0.197) and -0.0048 (0.036) per day for the IPSVT and IPSVC respectively indicate very minimal change over the 2 weeks of IPSV ratings.

## Discussion

The purpose of this study was to investigate if [1] clinicians could detect voice changes reported by teachers, by judging F0 and dB SAL data from IPSV tasks that were previously recorded and self-rated by teachers; and [2] the changes detected by the clinicians would be similar to those detected by the teachers themselves. The clinicians were able to use the F0 and dB SAL data from IPSV tasks to detect changes in the tasks using a scale of 1 – 10 (10 being the poorest), and to detect periods of what appeared to be vocal instability. However, the agreement between each set of teachers' self ratings (IPSVT) and the clinicians' ratings (IPSVC) were not statistically significant. These findings are consistent with those reported by Lee et al. (41), who found poor inter-rater agreement between patient and clinician ratings of perceived dysphonia.

Although the relationships between teacher and clinician ratings were not statistically significant, descriptive results (Table 2) offer some insights into the IPSV rating scale and its usefulness. First, the absolute mean *difference score* between the IPSVT and the IPSVC for all teachers was 1.7 (1.4). Thus, on the average, the teachers' and clinicians' ratings were within approximately 2 points of each other. Individual IPSVT standard deviations were below 2 for all but one teacher, indicating that the majority of teachers did not report a large variability in their ability to perform the IPSV task. This stability was further demonstrated by the small group mean slope for the IPSVT. These results are consistent with the slope analysis of IPSV in a larger group of teachers (42).



The range of self-reported mean IPSVT ratings (from 1.7 to 7.8) was much larger than the range of mean IPSVC ratings (from 2.3 to 4.9). This indicates that IPSVC ratings tended to be more conservative and contain a smaller range of numbers. In fact, mean IPSVC ratings were lower (better) than mean IPSVT ratings for 80% of the teachers. Lee et al. (41) also found that patients tended to rate their voices more severely than the clinicians did.

Differences between the IPSVT and IPSVC ratings could be an indication that the clinicians and the teachers may not have been using the same “anchor” for the “best” voice which could be produced. For example, if the teachers’ memory of “best” voice were from outside the 2 weeks of dosimetry recording, they would never give a “1” to a production over the 2 weeks of recording. However, because the clinicians only had the data from within the 2 weeks of recording upon which to base their judgment, they were more likely to give a “best” rating of “1”. This is especially evident with F072. The mode of the IPSVC for this teacher was 3, while the mode of the IPSVT was 9. Thus, F072 consistently rated herself on the higher (worse) end of the scale than the clinicians did. However, as can be seen, the standard deviation in ratings for both the IPSVT and IPSVC were small (1.6, 0.7) respectively.

Further evaluation of F072 offers insight into some of the merits of the IPSV for self rating. It is interesting to note that F072 had the highest VHI score of any of the teachers (a score of 58), and also the highest mean IPSVT score 7.8 (1.6). On a day that F072 reported that she felt she was “straining to talk”, she also commented that the IPSV tasks were difficult to do softly (correctly). Additional analysis of the two week dosimetry data revealed that F072 was phonating during 27% of the total recording time, which was the highest percentage of phonation time for any of the teachers in this paper. In addition, F072 had an average dB SAL level of 73, the fourth loudest of the group. It can be speculated that the dB SAL level, combined with the high percent of talk time, could create an environment of excessive tissue vibration and vocal fold swelling. This might then result in more difficulty completing the IPSV tasks, and thus higher IPSV ratings. This, in combination with the high VHI score for F072, is an indication of the possible sensitivity of the IPSV in the self identification of voice problems, and suggests external validity of the IPSV measure. These relationships should be further tested.

IPSVT and IPSVC ratings that were within 0 or 1 point of each other, and individual mean modes that were within one digit, indicate that the clinicians and teachers were detecting similar trends in voice performance. A comparison of the mode of the IPSVT and IPSVC for each teacher demonstrates that for 50% of the teachers, the teachers and clinicians had either the same mode, or were within one digit of each other for the most frequently occurring number.

In three specific cases (F080, M059 and F064), teacher and clinician ratings were fairly well matched. As can be seen in Table 2, IPSVT and IPSVC ratings for F080, M059 and F064 were within 0 or 1 point of each other 78, 76 and 67% of the time respectively.

A few reasons for the similarity between IPSVT and IPSVC ratings for F080 could be that both the IPSVT and IPSVC were rated within a narrow range of values (a small standard

deviation, 1.1 and 0.6 respectively; and a mode of 2 for both IPSVT and IPSVC). F080 also had a clean data set (low occurrence of noise and other anomalies), which could have made the IPSVC files easier to rate. In addition, F080 may have been more ‘in-tune’ to changes in her voice that were reflected in her IPSV performance. For example, F080 indicated that her “voice felt raw” on the day that she had her highest IPSVT score of a 6, with a corresponding IPSVC score of 3 (IPSVC scores did not go higher than 3 for this teacher). Thus, both the IPSVC and IPSVT reflected the difficulties that this teacher was feeling. This also suggests that the IPSVT was an indicator of vocal fatigue for F080 on this day.

Analysis of individual ratings for M059 revealed that, although the magnitude of change may not have been consistently the same between IPSVT and IPSVC ratings (resulting in agreement that was not significant), the IPSVT and IPSVC ratings follow a similar direction of change (Figure 4). This, together with the similarities in IPSVT and IPSVC ratings seen with F080, provide support for further development of the IPSV as a voice rating scale that could be used by clinicians.

In order to improve the utility of the IPSV as a clinical tool, it is important to consider why IPSVC and IPSVT ratings were more similar for some individuals than for others. In cases of fewer similarities, several factors could have influenced the IPSVT and IPSVC agreement score. The most significant of these is that when the teachers rated their self production of the IPSV, they were able to hear themselves. The clinicians, on the other hand, did not have access to full acoustic data of the teachers’ voice, as they only had a visual and synthesized acoustic representation of the dB SAL and F0. Thus, perceptually the clinicians could not hear voice quality changes such as breathiness or hoarseness, or difficulties with voice onset which might have influenced the teachers’ self ratings.

In addition to lack of access to the full acoustic signal, the clinicians also did not have information regarding the teachers’ self-perception of effort to produce phonation (both at the level of the larynx as well as more generally), or the amount of discomfort the teachers might have felt during the task. If a teacher was able to produce phonation softly and at a high pitch with no breaks, but it felt effortful or uncomfortable, this may have influenced the teacher’s rating in a negative direction. The clinicians would have only been able to see that the pitch and loudness met the appropriate criteria, and that there did not appear to be any voice breaks, thus they may have given the teacher a good rating. An example of this occurred during the 14<sup>th</sup> day of ratings for M045, where the IPSVT rating was an 8, while the IPSVC rating was a 1. When investigating the logbook notes for this day, M045 did not report anything unusual about his voice. However, this rating did occur on a Friday. We could hypothesize that M045’s voice was tired from the work week, which likely caused M045 to experience discomfort or felt that the rating took effort. However, while this data was collected, it was not made available to the clinicians during their rating, nor was it used in the analysis. The goal in excluding this data was to keep ratings of IPSV and effort and discomfort separate. Nevertheless, some teachers may not have completely separated these ratings, and effort and discomfort may have still played a role in their IPSV ratings. Future studies should include instructions to help the teachers separate ratings of effort and discomfort from their IPSV ratings.



Another limitation of the current study is that although the teachers were told to use their “soft, high phonation” they were not provided with a specific starting pitch or dB level to use as a reference for the IPSV tasks. This allowed teachers to make the task “easier” for themselves (either consciously or unconsciously) by lowering their pitch or getting louder. For the clinicians, if performance on an IPSV task seemed too low in pitch or too loud for a particular teacher (based on that teacher’s “gold standard” as explained above in *Clinician Experimental Procedures*), it was difficult to know exactly why softer, higher phonation was not used. Reasons could include actual difficulty with the task (inability to reach a certain pitch/loudness), the ability to reach a certain pitch or loudness but avoidance of these targets in order to maintain a stable voice, the ability to achieve the appropriate pitch or loudness but “forgetting” what the target was, or simply a lack of awareness that the intensity was louder or the pitch was lower than it should be. To standardize performance in future studies, teachers will be instructed to produce their softest (as soft as possible without whispering), highest phonation at a specific starting pitch for each task, even if it produces periods of instability or aphonia.

Another reason for poor agreement in this study could be that many points on the clinician’s 1–10 scale had specific definitions. For example, the definition for a two was, “Pitch and/or loudness is slightly worse than a 1, all tasks are completed correctly, may have slight unsteadiness, if aphonia is present it is likely due to soft voice or register shift”. The 1–10 scale used by the teachers was not as specific, i.e. the teachers were told to look for general characteristics, such as roughness/breathiness, times when their voice didn’t come out at all, when only loud voice came out, when it took a while for their voice to activate, or when the sound cut out as they used their voice. In addition, as teachers became familiar with the tasks, they may not have continued to refer to these criteria as they made their ratings.

Finally, the dosimeter itself may have contributed to rating inconsistencies between teachers and clinicians. For example, periods of what appeared to be aphonia or instabilities in the voice for some individuals could have actually been a result of a noisy signal, or instances where the signal was too close to the noise floor. It could have been difficult to set noise floor levels for some teachers who were able to produce very soft phonations. Future studies should include more extensive evaluation regarding the optimum setting for noise floor levels and frequent monitoring of the noise floor levels throughout the study.

The principle components of F0 (pitch) and dB SAL (intensity) measures did provide important information which resulted in IPSVT and IPSVC ratings within 0 or 1 point on many occasions. However, when IPSVT and IPSVC ratings were not within 0 or 1 point, it is logical to question if these two measures alone are sufficient. Previous studies cite the use of F0 and intensity to measure effects of vocal loading (16, 19); however, in these studies F0 and intensity were used as a part of a larger battery of acoustic measures. In addition, results from the current study suggest that factors such as vocal hyperfunction, which may contribute to a sense of vocal fatigue (15), as well as changes in voice quality, may also play an important role in rating of IPSV. Future studies should incorporate a full acoustic signal to provide more information for IPSVC ratings.

## Conclusion

This study has demonstrated the potential usefulness of the inability to produce soft voice (IPSV) rating as a simple tool to detect voice changes by self or others. While it was hypothesized that fundamental frequency and intensity would be most sensitive to the effects of vocal fatigue, and were therefore selected as the principle components for clinician rating of IPSV scores in this study, other components of the voice signal as mentioned in the discussion could have had an influence. Although the agreement between self and clinician ratings were not statistically significant, strong relationships were demonstrated in some subjects. This study has provided valuable information regarding how to better control extraneous variables to make the IPSV a more reliable and consistent measure. In fact, many of the changes suggested in this paper have already been incorporated into current data collection being conducted by the authors. Tools that enhance early detection of vocal changes are important in the implementation of injury prevention and recovery.

## Acknowledgments

This study was made possible by funding from NIH/NIDCD R01-DC004224

## References

1. Roy N, Merrill R, Thibeault S, Gray S, Smith E. Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res.* Apr.2004 47:281–293. [PubMed: 15157130]
2. Titze I, Lemke J, Montequin D. Populations in the U.S. workforce who rely on voice as a primary tool of trade: A preliminary report. *J Voice.* 1997; 11(3):254–259. [PubMed: 9297668]
3. Sliwinska-Kowalska M, Niebudek-Bogusz E, Fiszer M, Los-Spychalska T, Kotylo P, Sznurowska-Przygocka B, Modrzewska M. The prevalence and risk factors for occupation voice disorders in teachers. *Folia Phoniatr Logop.* 2006; 58:85–101. [PubMed: 16479131]
4. Simberg S, Sala E, Vehmas K, Laine A. Changes in the prevalence of vocal symptoms among teachers during a twelve-year period. *J Voice.* 2005; 19:95–102. [PubMed: 15766854]
5. Verdolini K, Ramig L. Review: Occupational risks for voice problems. *Logoped Phoniatr Vocol.* 2001; 26:37–46. [PubMed: 11432413]
6. Roy N, Merrill R, Thibeault S, Gray S, Smith E. Voice disorders in teachers and the general population: Effects on work performance, attendance, and future career choices. *J Speech Lang Hear Res.* Jun.2004 47:542–551. [PubMed: 15212567]
7. Smith E, Lemke J, Taylor M, Kirchner H, Hoffman H. Frequency of voice problems among teachers and other occupations. *J Voice.* 1998; 12(4):480–488. [PubMed: 9988035]
8. Smolander S, Huttunen K. Voice problems experienced by Finnish comprehensive school teachers and realization of occupational health care. *Logoped Phoniatr Vocol.* 2006; 31(4):166–17. [PubMed: 17114129]
9. Smith E, Gray S. Frequency and effects of teachers' voice problems. *J Voice.* 1997; 11:81–87. [PubMed: 9075180]
10. Thibeault S, Merrill R, Roy N, Gray S, Smith E. Occupational risk factors associated with voice disorders among teachers. *Ann Epidemiol.* Nov.2004 14:786–792. [PubMed: 15519901]
11. Lowell S, Barkmeier-Kraemer J, Hoit J, Story B. Respiratory and laryngeal function during spontaneous speaking in teachers with voice disorders. *J Speech Lang Hear Res.* 2008; 51:333–349. [PubMed: 18367681]
12. Gotaas C, Starr C. Vocal fatigue among teachers. *Folia Phoniatr Logop.* 1993; 45:120–129.
13. Welham N, Maclagan M. Vocal fatigue: current knowledge and future directions. *J Voice.* 2003; 17:21–30. [PubMed: 12705816]

14. Kitch J, Oates J. The perceptual features of vocal fatigue as self-reported by a group of actors and singers. *J Voice*. 1994; 8(3):207–214. [PubMed: 7987422]
15. Solomon N. Vocal fatigue and its relation to vocal hyperfunction. *Int J Speech Lang Pathol*. 2008; 10(4):254–266. [PubMed: 20840041]
16. Laukkanen AM, Ilomaki I, Leppanen K, Vilkmán E. Acoustic measures and self-reports of vocal fatigue by female teachers. *J Voice*. 2008; 22(3):283–289. [PubMed: 17134877]
17. Laukkanen AM, Jarvinen K, Artoski M, Waaramaa-Maki-Kumala T, Kankare E, Sippola S, Syrja T, Salo A. Changes in voice and subjective sensations during a 45-min vocal loading test in female subjects with vocal training. *Folia Phoniatr Logop*. 2004; 56(4):335–346. [PubMed: 15557772]
18. Lehto L, Laaksonen L, Vilkmán E, Alku P. Occupational voice complaints and objective acoustic measurements – do they correlate? *Logoped Phoniatr Vocol*. 2006; 31:147–152. [PubMed: 17114126]
19. Rantala L, Vilkmán E. Relationship between subjective voice complaints and acoustic parameters in female teachers' voices. *J Voice*. 1999; 13(4):484–495. [PubMed: 10622515]
20. Rantala L, Paavola L, Korkko P, Vilkmán E. Working-day effects on the spectral characteristics of teaching voice. *Folia Phoniatr Logop*. 1998; 50(4):205–211. [PubMed: 9819482]
21. Chang A, Karnell MP. Perceived phonatory effort and phonation threshold pressure across a prolonged voice loading task: A study of vocal fatigue. *J Voice*. 2004; 18(4):454–466. [PubMed: 15567047]
22. Solomon NP, DiMatta MS. Effects of a vocally fatiguing task and systemic hydration on phonation threshold pressure. *J Voice*. 2000; 14(3):341–362. [PubMed: 11021502]
23. Boucher VJ, Ahmarani C, Ayad T. Physiologic features of vocal fatigue: electromyographic spectral-compression in laryngeal muscles. *Laryngoscope*. 2006; 116(6):959–965. [PubMed: 16735908]
24. Bastian R, Keidar A, Verdolini-Marston K. Simple vocal tasks for detecting vocal fold swelling. *J Voice*. 1990; 4(2):172–183.
25. Hunter E, Titze I. Quantifying vocal fatigue recovery: Dynamic vocal recovery trajectories after a vocal loading task. *Ann Otol, Rhinol Laryngol*. In press.
26. Titze I. Toward occupational safety criteria for vocalization. *Logoped Phoniatr Vocol*. 1999; 24:49–54.
27. Gray S, Titze I. Histologic investigation of hyperphonated canine vocal cords. *Ann Otol Rhinol Laryngol*. 1988; 97(4 Pt 1):381–388. [PubMed: 3408113]
28. Titze I, Svec J, Popolo P. Vocal dose measures: quantifying accumulated vibration exposure in vocal fold tissues. *J Speech Lang Hear Res*. 2003; 46:919–932. [PubMed: 12959470]
29. McCabe D, Titze I. Chant therapy for treating vocal fatigue among public school teachers: a preliminary study. *Am J Speech Lang Pathol*. 2002; 11:356–369.
30. Carroll T, Nix J, Hunter E, Emerich K, Titze I, Abaza M. Objective measurement of vocal fatigue in classical singers: a vocal dosimetry pilot study. *Otolaryngol Head Neck Surg*. 2006; 135:595–602. [PubMed: 17011424]
31. Spielman J, Hunter E, Halpern A, Titze I. Measuring improvement in teachers with voice complaints using the inability to produce soft voice (IPSV) rating: preliminary data. *Am J Speech Lang Pathol*. In review.
32. Titze I, Hunter E, Svec J. Voicing and silence periods in daily and weekly vocalizations of teachers. *J Acoust Soc Am*. 2007; 121(1):469–478. [PubMed: 17297801]
33. Popolo P, Svec J, Titze I. Adaptation of a pocket PC for use as a wearable voice dosimeter. *J Speech Lang Hear Res*. 2005; 48:780–791. [PubMed: 16378473]
34. Sapir S, Mathers-Schmidt B, Larson G. Singers' and non-singers' vocal health, vocal behaviours, and attitudes towards voice and singing: indirect findings from a questionnaire. *Eur J Disord Commun*. 1996; 31:193–209. [PubMed: 8776440]
35. Jacobson B, Johnson A, Grywalski C, Silbergleit A, Jacobson G, Benninger M, Newman C. The Voice Handicap Index (VHI): Development and validation. *Am J Speech Lang Pathol*. 1997; 6(3): 66–70.

36. Svec J, Popolo P, Titze I. Measurement of vocal doses in speech: experimental procedure and signal processing. *Logoped Phoniatr Vocol.* 2003; 28(4):181–192. [PubMed: 14686546]
37. Svec J, Titze I, Popolo P. Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *J Acoust Soc Am.* 2005; 117(3, Pt.1):1386–1394. [PubMed: 15807026]
38. Karnell M, Melton S, Childes J, Coleman T, Dailey S, Hoffman H. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice.* 2007; 21(5):576–590. [PubMed: 16822648]
39. Fleiss, J., Levin, B., Cho Pak, M. *Statistical Methods for rates and proportions.* 3rd. Hoboken: John Wiley & Sons; 2003. p. 598-624.
40. Hunter EJ. A comparison of a child's fundamental frequencies in structured elicited vocalizations versus unstructured natural vocalizations: A case study. *Int J Pediatr Otorhinolaryngol.* 2009; doi: 10.1016/j.ijporl.2008.12.005
41. Lee M, Drinnan M, Carding P. The reliability and validity of patient self-rating of their own voice quality. *Clin Otolaryngol.* 2005; 30:357–361. [PubMed: 16209679]
42. Hunter, E. General Statistics of the NCVS self-administered vocal rating (SAVRa). NCVS on-line technical memo 11. [cited 2008 August 12] [9 screens]. Available from: URL: <http://www.ncvs.org/e-learning/technical.html>

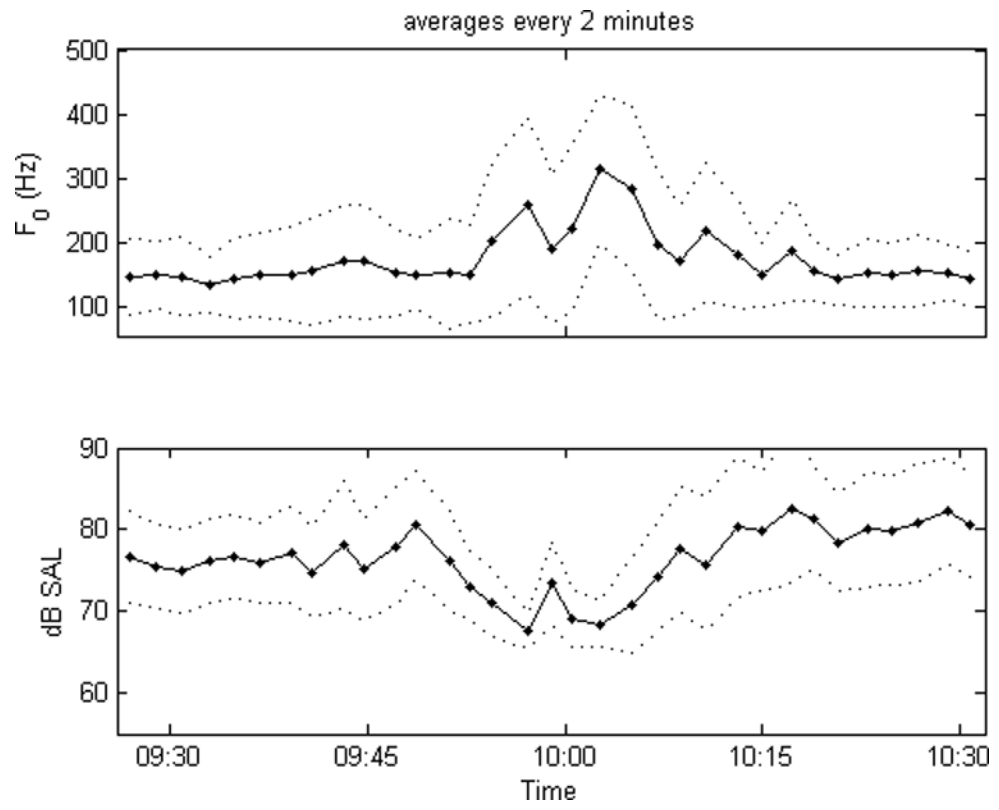


Fig. 1.

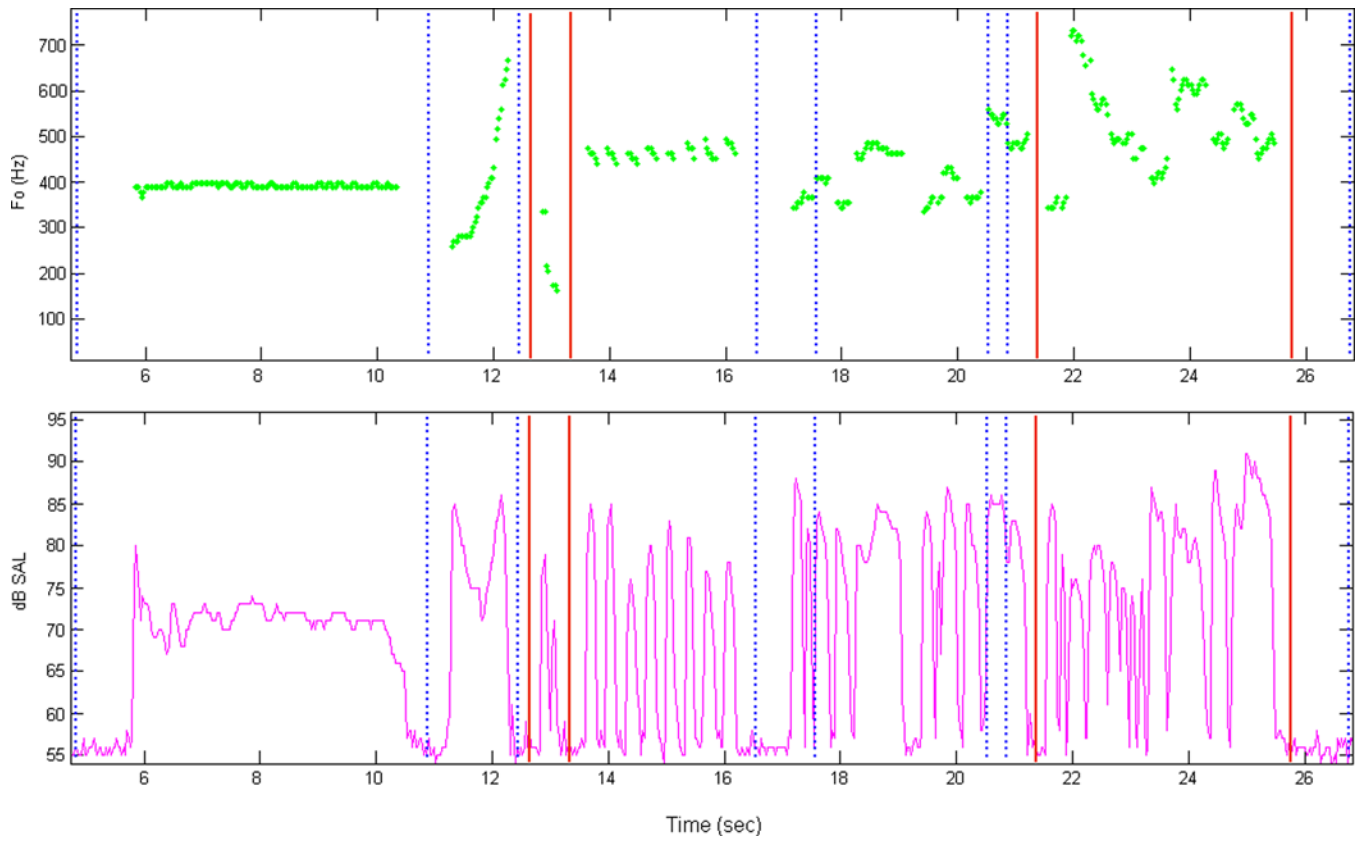


Fig. 2.



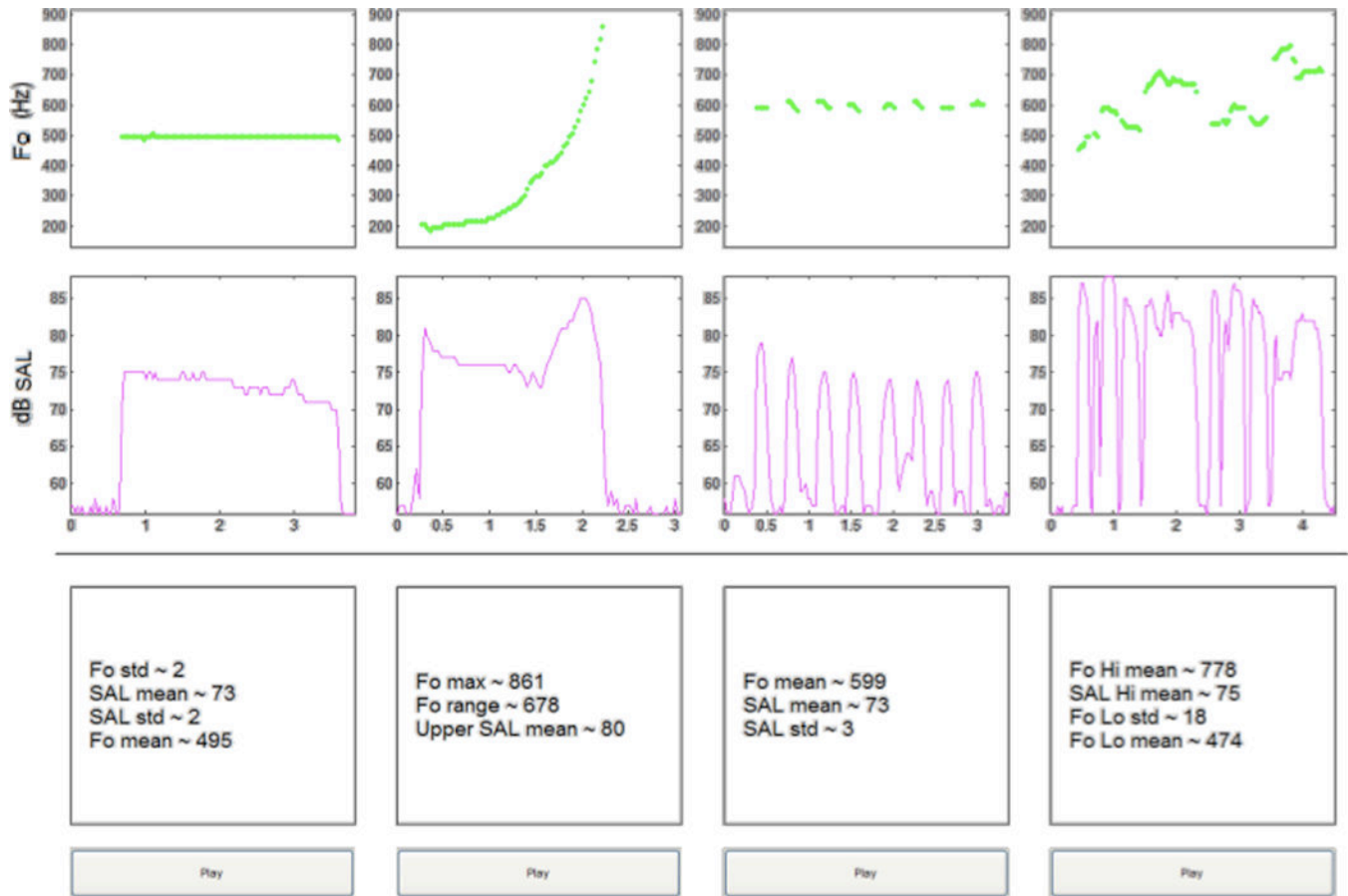


Fig. 3.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

### IPSVT and IPSVC for M059

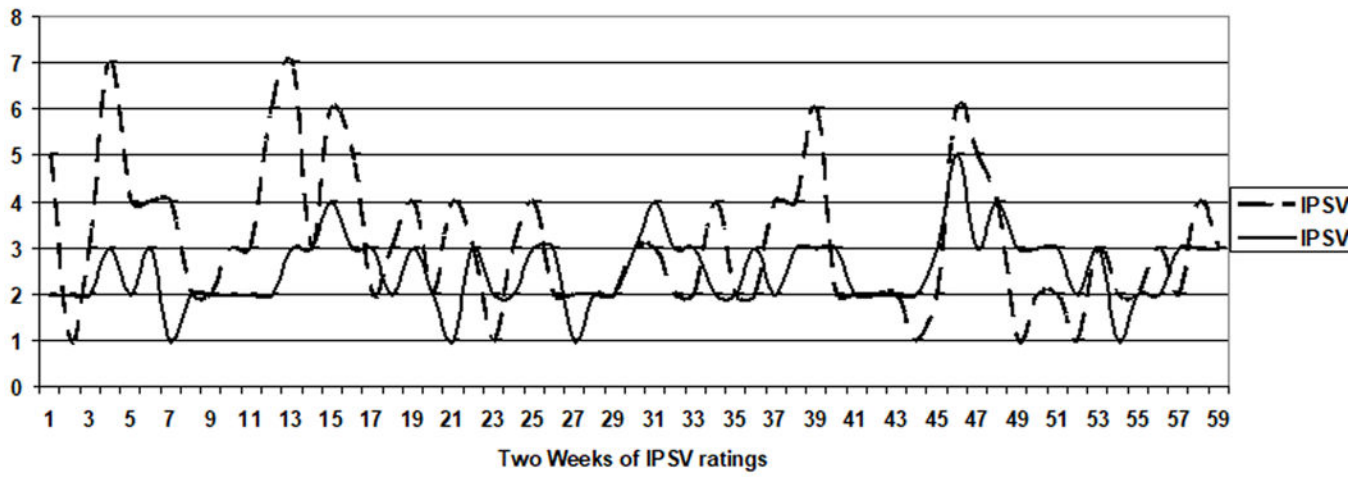


Fig. 4.

Table 1

Biographical and voice information regarding the study population (N=10).

Teacher F=Female M=Male	Percent Phonation Time	Age	VHI Total Score	VHI Self Rating	Self- reported Vocal Fatigue	Self-reported Voice Problems	Grade Taught	Subject Taught	Amplifier Worn
F064	15.8%	45	N/A	Normal	Yes	Hoarseness, sudden voice change, tightness	K-5	Phys. Ed.	No
F072	27.1%	48	58	Moderate	Yes	Hoarseness, sudden voice change, reduced pitch, discomfort, pain, tightness	4	General	Occasional
F079	9.7%	56	14	Mild	Yes	Reduced pitch	1-7	Phys. Ed.	No
F080	12.6%	29	18	Normal	Yes	Tightness	K-5	Music	No
F089	18.7%	31	22	Mild	Yes	Hoarseness, sudden voice change, discomfort	1-8	Music, Choir	Occasional
F104	19.2%	50	4	Normal	No	None	3-4	General	No
M045	18.7%	29	20	Normal	Yes	Hoarseness, sudden voice change, pain, discomfort	9-12	Theatre	No
M056	7.6%	53	33	Normal	No	Tightness, sudden voice change	High School	Physical Education	No
M057	19.2%	45	N/A	Normal	Yes	Hoarseness, discomfort, tightness, reduced pitch	9-12	Biology, Drama, Govt., Health	Yes
M059	15.1%	59	12	Normal	No	N/A	Elem., Middle School	English, Math, Principal	No

Note: Vocal Fatigue and “self-reported voice problems” were obtained from the baseline Vocal Health Questionnaire. VHI total score and self-rating were obtained from the baseline VHI. “Percent Phonation Time” is based on the percent of time phonation occurred while the dosimeter was attached and turned on.

**Table II**

Summary of results for IPSVT and IPSVC ratings for all subjects.

Teacher (# of files rated)	IPSVC vs. IPSVT					
	Mean IPSVT (SD)	Mean IPSVC (SD)	Avg. Diff (SD)	Mode of Avg. Dif. Score	% Agreement w/in 0 or 1	IPSVT Mode
F064 (105)	3.4 (1.3)	2.3 (0.7)	-1.1 (1.5)	0	67	3
F072 (86)	7.8 (1.6)	3.1 (0.7)	-4.7 (1.7)	-5	2	9
F079 (71)	3.7 (1.1)	3.3 (1.8)	-0.4 (1.9)	-2	49	4
F080 (86)	2.5 (1.1)	2.3 (0.6)	-0.2 (1.2)	-0.2	78	2
F089 (73)	1.7 (1.0)	4.5 (2.1)	2.7 (2.1)	1	37	1
F104 (65)	3.7 (1.6)	4.9 (2.0)	1.1 (2.3)	2	34	3
M045 (91)	5.3 (1.0)	2.6 (0.8)	-2.7 (1.5)	-2	23	5
M056 (75)	5.3 (3.5)	3.6 (1.2)	-1.8 (3.5)	1	31	2
M057 (81)	4.5 (1.8)	2.5 (0.7)	-2.0 (1.9)	-1	48	3
M059 (59)	3.1 (1.5)	2.5 (0.8)	-0.6 (1.5)	0	76	2
Group Mean (SD)	4.1 (1.5)	3.1 (1.1)	-1.0 (2.0)	-0.6 (2.0)	44.5 (24.2)	3.4 (2.3)
* Abs Mean (SD)			1.7 (1.4)	1.4 (1.5)		2.9 (0.9)

\* Absolute Group Mean and Standard Deviation

Note: Results of the Cohen's kappa were non-significant ( $p > 0.05$ ).