



## THE EFFECTS OF A REMEDIAL SINGING METHOD ON THE VOCAL PITCH ACCURACY OF INACCURATE ELEMENTARY SINGERS

**Karen A. Miyamoto**

University of Hawaii, Manoa

[miyamotok001@hawaii.rr.com](mailto:miyamotok001@hawaii.rr.com)

### INTRODUCTION

Singing is an important skill to be developed in the elementary music classroom. The National Standards for Music Education developed by The Music Educators National Conference (1994) specify that all children should be taught to sing (p. 26).

The condition of inaccurate singing has been found to be a detriment. Greenberg (1970), in discussing the effects of inaccurate singing, concluded that a child that knows something is wrong with his or her singing withdraws from most phases of the music program. Yuba (1998) suggested that inaccurate singing may result in embarrassment, humiliation, and even societal maladjustment.

Statistics indicate that inaccurate singing is an ongoing problem. The National Assessment of Educational Progress for the years 1971-1972, indicated that 50% of the nine-year-olds, 45% of the thirteen-year-olds, 35% of the 17-year-olds, and 30% of the adults were unable to sing the song *America* with acceptable pitch. According to Roberts and Davies (1975), approximately 18% of children in grade six and below were unable to sing simple songs in tune and were considered to be inaccurate singers. Goetze (1985) reported that 70% of the kindergarten, first- and third-grade subjects in her study were inaccurate singers. Aaron (1991) reported that 69% of the fourth, fifth, and sixth graders in his study were inaccurate singers.

Remedial techniques to correct inaccurate singing need to be investigated in order to diminish the dilemma of inaccurate singing. Gordon (1985) concluded that music educators used empirically unproven remedial techniques to work with inaccurate singers. He also found the converse to be true—that empirically proven techniques were not used or known to music educators. Gordon found that music educators often simply increased the amount of time spent on singing in an attempt to improve singing accuracy. He saw this as a sign of frustration by music educators who knew no other empirically corroborated means to adequately deal with singing inaccuracy.

Extant research indicates that there may be a multitude of causes and conditions for singing inaccuracy. This suggests that a multitude of vocal remediation strategies may be necessary in correcting inaccurate singing. Some causes, as indicated by previous research include piano accompaniment (Clayton, 1986), music aptitude (De Yarman, 1972; Jones, 1993; Jaffurs, 2000), development and maturation (Wilson, 1970; Levinowitz, Barnes, Guerrini, Clement, D'April, and Morey, 1998), singing range (Cleall, 1970; Guerrini, 2002),

singing an entire phrase versus individual melodic items (Petzold, 1966), age (Mizener, 1993), pitch discrimination (Bentley, 1968; Zwissler, 1972), self concept (Greenberg, 1970), unison versus individual singing (Goetze 1985; Smale 1987), singing with text versus singing on a neutral syllable (Goetze, 1985; Flowers and Dunne-Sousa, 1990), home environment and heritage (Eikum, 1963), the vocal model (Yarbrough, Bowers, and Benson, 1992; Sims, Moore, and Kuhn, 1982; Small and McCachern, 1983; Montgomery, 1988; Green, 1990; Gratton, 1992), and vocal registration (Brown, 1988).

It is relevant to review the existing research on remedial singing techniques to build upon extant information. Additional drills and exercises have been employed as remedial treatment. Roberts and Davies (1975) successfully utilized speech devices to extend the speaking range and exercises in finding a personal note. They were able “to effect some improvement in pitch production among children rated by their teachers as monotones” (p. 236). Richner (1976) found that remedial voice instruction had a significant positive effect on the ability of inaccurate singers to reproduce pitches. Roberts and Davies (1976) investigated vocal range extension. The results indicated that the remedial group showed greater improvements on single note production and interval production. Buckton (1977) found that a vocal program significantly improved the vocal accuracy scores of the vocal group over instrumental and control groups. Rooks (1987) investigated the effects of remedial vocal training on inaccurate singers. Findings indicate that restricted range singers trained in both the high and low range gained significantly more accuracy than those trained in only the high range.

Several studies focused on psychological aspects to aid the inaccurate singer. Jones (1971) investigated the use of a vertically arranged keyboard instrument. The visual representation of the vertical keyboard as it related to “high” and “low” pitches helped the inaccurate singer. Jarjisian (1981) found that young children’s rote singing achievement was benefited by pitch pattern instruction, which included both diatonic and pentatonic patterns. Apfelstadt (1984) investigated the effects of melodic perception instruction. She found significant differences on vocal pitch-pattern accuracy and in rote singing accuracy. Kramer (1985)--found that imagery training improved the ability of inaccurate singers to match pitches vocally. Welch, Howard, and Rush (1989) explored the use of visual feedback using a microcomputer. Subjects improved over the treatment period, and it was concluded that “verbal feedback on its own appears to be less powerful in promoting learning than real-time, meaningful visual feedback” (p. 156). Matthias (1997) investigated the use of sequential games. Vocal accuracy was said to have improved after completing a sequential series of pitch matching games.

Extant physiologically based studies have focused primarily on the area of breath control management. Phillips (1983) found that breath control management significantly improved vocal range, vocal intensity, vocal duration, and pitch accuracy. Gackle (1987) examined the effects of selected vocalizes that employed breath management techniques. She found that the exercises significantly improved pitch perturbation. Aaron (1991) found that vocal coordination instruction was more effective in improving vocal pitch accuracy for boys than for girls and that highly inaccurate boy singers benefited the most from vocal coordination instruction. Phillips and Aitchison (1995) found that vocal range was improved through breath control management instruction. Collins (2000) concluded that breath management may be so interdependent with students' abilities to coordinate the vocal mechanism that it alone may not produce significant results in vocal performance.

Research in the area of physiological perspectives with regard to singing, includes that of Yuba (1998), who developed a vocal training method that was intended to specifically train the cricothyroid muscle, which is used while singing (Yuba 2001, p. 1). Based upon his theory of cricothyroid muscle function, Yuba (2002) explained that mechanically, the cricothyroid muscle determines the pitch like a guitar reel or spool. He said that its main function is to act as tensors, tilting the thyroid cartilage forward and

downward, lengthening the vocal folds and making them thinner, resulting in raised pitch. Yuba added that conversely, its relaxation lowers the pitch. He elaborated that the preponderance of the cricothyroid muscle against the closing muscle group, or arytenoid muscles, produces a head voice register made up of breathy sounds because it cannot close the glottis completely. Yuba added that on the other hand, the preponderance of the closing muscle group, or arytenoid muscles, against the cricothyroid muscle, produces a chest voice register, or non-breathy sounds (p. 4).

Figure 1 provides a view of the intrinsic muscles of the larynx.

Yuba explained that the quality of vibration of the vocal folds is determined by the balance of three factors:

1. the action of the cricothyroid muscle stretching the vocal folds which mainly elevates the pitch;
2. that of the closing muscle group (lateral cricoarytenoid muscle, transverse arytenoid muscle, and oblique arytenoid muscle), which closes the glottis along with the pressure of expiration; and
3. the physical movement of articulation (Yuba 2000, p. 2).

Yuba (2000) said that a subsidiary result of the method was to correct inaccurate singing (p. 2). He noted that to date, he has corrected over 900 inaccurate singers (\_\_\_\_, 2003).

Yuba (2002) devised a series of musical exercises based on his philosophy. Following are the Yuba Method basic steps in correcting inaccurate singing:

1. distinguish between the head voice and the chest voice;
2. sing some very simple songs in the head voice and the chest voice;
3. sing from the head voice to the chest voice and then sing from the chest voice to the head voice (p. 4).

### **DELIMITATIONS AND PURPOSE OF THE STUDY**

The Yuba Method appears to have had some success in correcting inaccurate singers although there is no existing empirical data to support it. Furthermore, Yuba has not provided adequate evidence (e.g. via laryngoscopy) that indicates his exercises actually target training of the cricothyroid muscle in singing. It is not within the scope of this study to verify cricothyroid muscle functioning while utilizing the Yuba Method exercises. Rather, the purpose of this investigation was to determine the effects of the Yuba Method on the vocal pitch accuracy of inaccurate elementary school singers in grades four, five, and six.

### **RESEARCH QUESTIONS**

1. Will the remedial singing treatment significantly improve inaccurate singers over that of a control group?
2. How will the treatment affect “high,” “middle,” and “low” inaccurate singers?

### **METHOD**

This study used a pretest posttest control group design using the Yuba Method with a treatment group, and no remedial treatment with the control group to determine the effects of the Yuba Method. The subjects were fourth-, fifth-, and sixth-grade students (N = 320) in one public urban elementary school in Honolulu, Hawaii. This group comprised the total enrollment of these grades in the school, and consisted of 165 boys and 155 girls. The fourth grade was comprised of 51 boys and 53 girls. The fifth grade was comprised of 64 boys and

49 girls, and the sixth grade was comprised of 50 boys and 53 girls. The population of fourth, fifth, and sixth graders included all subjects between the ages of 8.5 to 9.4, 9.5 to 10.4, and 10.5 to 11.4 years of age respectively, by September 1, 2002. The ethnic make up of the school population consisted primarily of students of Asian and Pacific Islander heritage. The socio-economic status of the school consisted of a population of 40%, on free or reduced lunch.

Materials and equipment used in the study consisted of a Gateway laptop computer, an Electro Voice Microphone, a Yamaha PSR-540 Electronic Keyboard, a two foot by four foot mirror, a Sony CFD-V17 CD Player, and a Sony Hi-8 Video Recorder. The Sona-Speech Model 3600 software program by Kay Elemetrics was used to analyze the criterion pitches in the *Pretest Singing Stimulus* and the *Posttest Singing Stimulus* utilized in the study. The Sona-Speech Model 3600 software program was ordered from Kay Elemetrics Corporation at 2 Bridgewater Lane, Lincoln Park, New Jersey, 07035, USA. The Sona-Speech Model 3600 software program is the software-only component of the Visi-Pitch hardware device used in previous research (Goetze, 1985; Clayton, 1986; Smale 1987; Brown, 1988; Goetze & Horii, 1989; Moore, 1991). The Sona-Speech Model 3600 software program is a clinical package of speech training and analysis programs. The specific program utilized in the Sona-Speech was Real Time Pitch, which calculates frequency in Hertz of recorded pitches. The Yuba Method was obtained from *Dream Voice Training: Muscles For Singing* Tokyo, Japan: Victor Entertainment, Inc.

## PROCEDURE

The total enrollment of fourth-, fifth-, and sixth-grade students was taught to sing the Pretest Singing Stimulus commencing the first week of October 2002. The students were taught using the typical rote method of teaching in their regular general music class instructional period, which met once a week for 55 minutes. Twenty minutes of each subsequent class time was spent teaching the song stimulus. A criterion of 75% was established as a minimum attendance for participation in the study. No students were eliminated on this basis.

The *Pretest Singing Stimulus* was individually administered to the 320 fourth-, fifth-, and sixth-grade students during the first two weeks of November 2002. The *Pretest Singing Stimulus* was designed by the investigator to classify subjects either as “accurate” or “inaccurate” singers. The *Pretest Singing Stimulus* consisted of the first phrase of the Israeli folk song *Shalom Chaverim* in the key of D minor (Figure 2). During the administration of the *Pretest Singing Stimulus*, the first two starting pitches of the phrase were played on an electronic keyboard and subjects were required to sing the entire phrase “a cappella.” Audio Examples [1](#), [2](#), [3](#), [4](#), and [5](#) demonstrate VPA scores on the *Pretest Singing Stimulus* of 14, 48, 115, 289 and 360 respectively.

The Pretest Singing Stimulus test was analyzed and scored. Selected criterion pitches (D4, D5, C5, F4) were used to calculate singing accuracy rather than using a subject’s deviation on all of the notes in the test stimulus. This was based on previous research findings (Goetze, 1985, p. 75), which indicated that an average of selected notes was more descriptive of a subject’s singing accuracy than an average of all of the notes sung in a test stimulus.

It was desirable for the purposes of the present study to have the singing stimuli encompass both the chest and head registers since previous research suggests that the vocal register break is a possible cause of singing inaccuracy. Cooper (1995) found that children who had not yet learned to use the head voice register had difficulty matching pitches above the voice break (p. 36), and Guerrini (2002) found that once students were able to sing one song accurately using notes above the register break, they appeared to be able to sing other

songs accurately (p. 56).

The vocal register break has been examined in previous research. Cooper (1995) identified the break between the chest and the head voice to occur around G4 or A4 (p. 36). Phillips (1996) noted that the pitch F#4 was where there was a balance between the chest and head voices. The range of the *Pretest Singing Stimulus* in the present study was thus from A3 to D5 to encompass both the head and chest registers.

A vocal pitch accuracy score (VPA) was obtained for each student on the basis of the *Pretest Singing Stimulus*. The Sona-Speech Model 3600 software program was used to calculate the score. The vocal pitch accuracy score (VPA) was the average cent deviation of the four criterion pitches in the singing stimulus. The Sona-Speech software program was used to calculate in Hertz, the frequency of each selected criterion pitch of the *Pretest Singing Stimulus*.

The Sona-Speech sampled the recorded voice and displayed the frequency curve of the criterion patterns on the computer monitor. The investigator then moved cursors to outline the segments of the curve representing the pitch to be analyzed, and the Sona-Speech automatically calculated the frequency, in Hertz of the pitch area between cursors. Because frequencies in Hertz are not equal-interval data, logarithms were used to calculate the interval or deviation in cents, where 100 equal cents equaled one semitone between each response pitch and its corresponding stimulus. Calculation of the size of pitch intervals followed the Campbell and Greated procedure (1987, p. 77). The total deviation in cents between each response pitch and its corresponding stimulus was calculated.

Absolute values were used in these calculations to avoid the possibility of both positive and negative cent deviations. For example, sharp and flat responses, respectively, on different pitches within the pattern might cancel each other out. Therefore, although VPA scores represented overall deviation from the model or actual pitches, they did not provide an indication of the direction of deviation or contour of the response. Because VPA scores represented divergence from the model, lower scores indicated more accurate performance and higher scores indicated more inaccurate performance.

The total enrollment of fourth-, fifth-, and sixth-grade students (N=320) was classified into two groups—accurate (N = 152) and inaccurate singers (N = 168), based on the *Pretest Singing Stimulus* vocal pitch accuracy score (VPA). Subjects with a VPA score of 100 or greater were identified as inaccurate singers. Subjects with a VPA score below 100 were identified as accurate singers. The accurate singers were eliminated from the remainder of the study. The criteria of using the VPA score of 100 or greater to determine the inaccurate singer was used by Goetze (1986), Smale (1987), and Cooper (1995).

The formation of three subgroups utilizing the inaccurate singer VPA scores was the next step. All of the inaccurate singer scores (N = 168) were listed in ascending order from the lowest to the highest scores. The inaccurate singers were divided into three subgroups of equal size (N = 56) to constitute "low," "middle," and "high" VPA scores.

The Komolgorov-Smirnov test for comparing two populations was calculated between the three subgroups of "low," "middle," and "high" to determine if in fact the populations were different. Results of the Komolgorov-Smirnov test indicated at the  $p < .05$  confidence level that the three subgroups were from different populations. The three subgroupings were therefore deemed an appropriate design for the experiment.

Twenty subjects were randomly selected, from each of the three subgroups of "low," "middle," and "high" VPA inaccurate singers, to be either in the treatment group (N = 30) (Yuba Method) or control group (N = 30). There was no differentiation of gender or grade

level in this process.

Each subject in the study was assigned a five-digit subject number, which indicated the following: (a) Digit one represented the subject's grade level; (b) Digit two represented the group assignment. The group receiving the Yuba Method treatment was assigned the number "1." The control group was assigned number 2; (c) Digit three represented the "low" (1), "middle" (2), or "high" (3) groupings within the experimental or control groups; (d) Digits four and five represented the subject number within the treatment or control groups. For example, subject number 62101 was a sixth grader, the first subject in the control "low" group.

The total duration of the testing and treatment portions of the study was twelve weeks from the *Pretest Singing Stimulus* administration to the *Posttest Singing Stimulus* administration. The time period to complete the experimental treatment on all 30 treatment subjects lasted no longer than three weeks.

Each subject in the treatment group received one individual, 45-minute treatment session using the Yuba Method in addition to their regular music class, which occurred once a week for 55 minutes. Each subject in the control group received only instruction in their regular music class, which was the same as that of the treatment group. The regular music class lessons for that semester included singing with no remedial provisions, note reading, playing of the recorder, and audiation exercises. The control group received no instruction other than their regular music class.

All subjects in grades four, five, and six in the school were taught to sing the *Posttest Singing Stimulus* commencing one month prior to the testing and for a period of four consecutive weeks thereafter for a period of 20 minutes at each session. This occurred in their regular music class, which consisted of a heterogeneous grouping of accurate, control, and treatment singers.

The researcher chose two different songs for the singing stimuli--*Shalom Chaverim*, first phrase, and *The Star-Spangled Banner*, first phrase (Figure 3). Two different singing stimuli were chosen due to past research findings that mistakes were often carried over in the same song regardless of training (Goetze, 1985).

The criterion pitches selected for this study were D4, F4, C5, and D5 for both test stimuli (where middle C is C4). The criterion pitches were selected by the investigator in an attempt to span the tones across the vocal register break (D4-D5), a tone below the register break (D4), and a tone around the register break (F4). A song phrase was used rather than utilizing the matching of single tones because past research indicated that the matching of single pitches had no correlation to singing a song in tune (Flowers and Dunne-Sousa, 1990, p. 111).

The pretest and posttest stimuli for the present study were sung on the neutral syllable "loo." Previous research indicates that students sing more accurately on a neutral syllable (Gould, 1969; Goetze, 1985), and Edwin Gordon (1984) recommends that "students must echo in solo with a neutral syllable" (p. 30). Gordon also added that "the use of words of a song actually inhibits the learning of tonal syntax" (p. 143). As well, the neutral syllable, "loo" was used for the singing stimulus by Goetze, (1985), Smale (1987), and Cooper (1995).

The Yuba Method was administered as the remedial singing method (Yuba, 1998). The exercises were recorded on an audio CD and consisted of a female soprano singer as the vocal model over a synthesized instrumental accompaniment. [Audio Example 6](#) demonstrates Audio Track 12 as used in the Treatment Script (see Appendix). Subjects were

to echo the vocal model. Instructions were read from a script by the researcher, who also served as the treatment instructor. Subjects in the control group took the *Posttest Singing Stimulus* at the end of the three-week treatment period of the treatment subjects following their regular music class session. The subjects in the treatment group took the *Posttest Singing Stimulus* immediately following their individualized, 45-minute vocal training session, which consisted of the Yuba Method exercises.

Analysis of Variance (ANOVA) was employed to determine whether or not the treatment improved the singing ability of treatment subjects. The design consisted of a two-way classification, with the sources of variation being (1) the effect of the Yuba Method training, and (2) the pretest ranking of inaccurate subjects into the “low,” “middle,” and “high” subgroups. The analysis also provided an assessment of the variation contributed by the interaction of “main effects” (1) and (2) defined above.

## RESULTS

**Table 1.**

Percentage and Number of Accurate and Inaccurate Singers by Grade Level and Gender Based on *Pretest Singing Stimulus* VPA Scores.

Grade (N), Gender (N)	Accurate singers % (N)	Inaccurate singers % (N)
Grade 4 (104)	40.39 (42)	59.61 (62)
Boys (51)	39.22 (20)	60.78 (31)
Girls (53)	41.51 (22)	58.49 (31)
Grade 5 (113)	47.79 (54)	52.21 (59)
Boys (64)	45.31 (29)	54.68 (35)
Girls (49)	51.02 (25)	48.97 (24)
Grade 6 (103)	54.37 (56)	45.63 (47)
Boys (50)	71.00 (23)	24.27 (27)
Girls (53)	62.26 (33)	19.41 (20)
All (320)	47.50 (152)	52.50 (168)
Boys (165)	43.64 (72)	56.36 (93)
Girls (155)	51.61 (80)	48.38 (75)

Results of the *Pretest Singing Stimulus* are summarized in Table 1 by grade level and gender. The researcher-designed *Posttest Singing Stimulus* results yielded the data that represented the mean number of cents that subjects deviated from all four of the selected criterion pitches. High scores indicated highly inaccurate singing, and low scores indicated more accurate singing. The mean scores for each subgroup are provided in Table 2.

**Table 2.**

Mean *Posttest Singing Stimulus* VPA Scores between Groups and Subgroups

	Low	Middle	High
Treatment	76	109	85
Control	118	185	302

In order to determine the effects of the Yuba Method, subjects' *Pretest Singing Stimulus* VPA scores were compared with their corresponding *Posttest Singing Stimulus* VPA scores. Gain scores were computed by subtracting a subject's *Posttest Singing Stimulus* VPA score from the corresponding *Pretest Singing Stimulus* VPA score. Positive gain scores represented an increase in singing accuracy. Negative gain scores represented a decrease in singing accuracy. Pretest-posttest gain scores were computed for the treatment and control groups and a double classification analysis of variance was computed using gain scores to determine if there was a significant difference in the performance of subjects in the treatment group versus the control group ( $p < .05$ ).

Table 3 provides the Mean *Posttest Singing Stimulus* VPA gain scores between groups and subgroups.

	Low	Middle	High
	(Most Accurate)		(Least Accurate)
Treatment	+51.27	+ 82.91	+ 312.52
Control	+ 2.30	+ 7.32	+ 64.26

Three subjects were unable to produce the *Posttest Singing Stimulus* in a manner that could be reliably scored and so were dropped from the remainder of the study. Both the Shapiro-Wilk and the Kolmogorov-Smirnov tests of normality confirmed this and indicated that the raw scores, VPA, were non-normally distributed at  $p < .0001$  and  $p = .01$ , respectively. These were subjects 52102, a fifth-grade subject in the "low" control group, 41210, a fourth-grade subject in the "middle" treatment group, and 62308, a sixth-grade subject in the "high" control group.

In analyzing the gain scores, some of the VPA gain scores turned out to be negative numbers due to a decrease in singing accuracy on the *Posttest Singing Stimulus*. To compensate for this, 300 cents were added to all of the VPA gain scores for the calculations. Moore and McCabe (2003) explained that converting numerical descriptions of a distribution from one unit of measurement to another is a linear transformation of the measurements (p. 51). They explained that linear transformations do not change the shape of a distribution (p. 53).

In order to determine whether or not to employ parametric or nonparametric statistical procedures, tests for normality of the sampled population were calculated. Rainbow and Froehlich (1987) stated that parametric statistics are more powerful than nonparametric tests. They defined "powerful" in statistical terms to mean that a test discriminates between two sets of data in such a way that the null hypothesis may be rejected even if the differences in scores are relatively small. They further explained that because researchers are concerned about minimizing the probability that a null hypothesis is maintained when it is in fact false, the more powerful statistic should be given preference. Rainbow and Froehlich concluded that when presented with the choice of using parametric versus nonparametric tests in a research situation, parametric tests should be employed (p. 256).

The tests of normality were important because otherwise the statements about the probability were not likely to be true. Moore and McCabe (2003) explained that the decision to describe a distribution by a normal model determines the later steps in the analysis of the data (78).

The scores were tested for normality by the *SAS Statistical Software Program*. The residuals from the *Analysis of Variance* were normal. That is, both the *Shapiro-Wilk* and the *Kolmogorov-Smirnov* tests were non-significant at the  $p = .08$  and  $p > 0.15$  levels, respectively. Based on the aforementioned results, parametric statistics were employed for the analysis. The ANOVA was calculated by the General Linear Model Procedure (GLM) of SAS which is able to accommodate unequal numbers in experimental groups without introducing error in the probabilities.

**Table 4.**

The General Linear Model ANOVA Results for Log-Transformed VPA Scores with 300 Cents Added (N = 57)

Source	<i>df</i>	SS	Mean Square	<i>F</i>	<i>p</i>
Treatment	1	1.14	1.14	21.14	<.0001
Group	2	1.29	0.64	11.90	<.0001
Group x Treatment	2	0.43	0.22	4.01	0.0024

Table 4 presents the ANOVA *General Linear Model Procedure* of the log-transformed scores with 300 cents added to each score and 57 observations.

**Table 5.**

VPA Log-Transformed Gain Scores by Group.

Level of Treatment	N	Mean	SD
Treatment (1)	29	6.09	0.27
Control (2)	28	5.81	0.30

Table 5 presents the VPA log-transformed gain scores by level of treatment with 300 cents added to each score, and 57 observations.

**Table 6.**

Log-Transformed VPA Gain Mean Scores by Group Level.

Level of Group	N	Mean	SD
High	19	6.07	0.38
Middle	19	5.84	0.28
Low	19	5.84	0.11

Table 6 presents the data of the log-transformed VPA gain mean scores by level of group with 300 cents added to each score and 57 observations.

**Table 7.**

General Linear Model Procedure on Log-Transformed Scores with 300 Cents Added by Group.

Level of Group	N	Mean	SD

High Treatment	10	6.39	0.23
High Control	9	5.92	0.38
Middle Treatment	9	6.01	0.09
Middle Control	10	5.69	0.31
Low Treatment	10	5.86	0.10
Low Control	9	5.82	0.13

Table 7 presents the data of the log-transformed general linear model procedure by level of group and level of treatment with 300 cents added to each individual gain score and 57 observations. The equal values of the “middle” subgroup mean and the “low” subgroup mean produced unequal values for the back-transformed VPA gain scores in Table 8 due to the unequal SD values.

**Table 8.**

Back-Transformed VPA Gain Mean Scores by Level of Treatment Measured in Cents.

Level of Treatment	N	Mean
Treatment	29	157.24
Control	28	46.82

**Table 9.**

Back-Transformed VPA Gain Mean Scores by Level of Group Measured in Cents.

Level of Treatment	N	Mean
High	19	212.65
Middle	19	57.95
Low	19	45.61

The “log e” transformed differences were then back-transformed (Table 9) to make the data meaningful in terms of comparisons in cents using the formula by Haan (1977, p. 107). Tables 10 through 12 provide the back transformed mean scores for the General Linear Model log-transformed scores by group with 300 cents added to each score and 57 observations.

**Table 10.**

Back-Transformed VPA Gain Mean Scores by Group Measured in Cents.

Group	N	Mean
High Treatment	10	309.66
High Control	9	100.45
Middle Treatment	9	109.12
Middle Control	10	10.16
Low Treatment	10	52.31

Low Control	9	38.36
-------------	---	-------

Table 10 presents the VPA gain scores by subgroup, log-transformed and back-transformed as a means for comparison.

<b>Table 11.</b> VPA Gain Scores by Group, Log-Transformed and Back-Transformed.			
Level of Treatment	N	Log-Transformed Mean	Back-Transformed Mean
Treatment (1)	29	6.09	157.24
Control (2)	28	5.81	46.82

Table 11 presents the VPA gain mean scores by group level, log-transformed and back-transformed as a means for comparison.

<b>Table 12.</b> VPA Gain Mean Scores by Group Level, Log-Transformed and Back-Transformed.			
Level of Treatment	N	Log-Transformed Mean	Back-Transformed Mean
High	19	6.17	212.65
Middle	19	5.84	57.95
Low	19	5.84	45.61

Table 12 presents the general linear model procedure on log-transformed scores and back-transformed scores as a means for comparison by level of group. Table 13 provides the General Linear Model Procedure of the log-transformed and back-transformed scores by group and level.

The main effect of treatment was found to be highly significant at the  $p < .0001$  significance level. In addition, some subgroups benefited from the treatment more than others ( $p < .0024$ ). The treatment "high" subgroup (the most inaccurate singers) benefited the most, followed by the treatment "middle" subgroup, and lastly the treatment "low" subgroup.

## DISCUSSION

The results of this study indicate that the treatment, which consisted of the Yuba Method exercises, was highly effective in improving the vocal pitch accuracy of inaccurate elementary singers ( $p < .0001$ ). The treatment was also found to be most effective with highly inaccurate singers ( $p < .0024$ ). An example of an improved posttest VPA score by subject 61310, a subject in the treatment high subgroup, is demonstrated in [Audio Examples 7 & 8](#) (*Pretest Singing Stimulus VPA 756*, and *Posttest Singing Stimulus VPA 25*, respectively). Based on the results, the null hypothesis was rejected at the  $p < .0001$  significance level.

The investigator concedes that certain conditions may have compromised the interpretation of the results. As a result, some degree of caution should be maintained by the reader. These conditions follow:

1. The school population in this study may be unique and accordingly, the results might not be generalizable to the general population.
2. The two singing test stimuli were not tested for equitability of difficulty level. The *Pretest Singing Stimulus* was in a minor key and the *Posttest Singing Stimulus* was in a major key. This might have presented unequal levels of difficulty between the two test stimuli.
3. The treatment group had the advantage of additional instruction, which was lacking in the control group. Improvement in singing accuracy might be attributed to additional instruction and not necessarily to the Yuba Method.
4. The criteria for determining the accurate and inaccurate singer as used in this study, has not been validated by research. The VPA of 100 cents or greater to define the inaccurate singer, was arbitrarily selected by Goetze (1985) and needs to be confirmed through empirical research. Cooper (1995) recommended that “a study comparing subjective ratings of perceived accuracy with objective electronic accuracy evaluations of the same sample” be conducted (p. 230). Implications of these findings are that this criteria for determining inaccurate singer might thus have been inadequate and some of the singers in the study might thus have been accurate singers.
5. Most subjects, in both the control and treatment groups performed better on the *Posttest Singing Stimulus* than on the *Pretest Singing Stimulus*. This may have been due to decreased test anxiety, familiarity with the testing situation, or the possibility that the *Posttest Singing Stimulus* was a more familiar song. This song was also in a major key, which may have made it easier to sing as opposed to a song in a minor key.

Following are recommendations for future research based on the results of this study:

1. Research studies should be conducted to determine more precise differences between the echo singing of phrases versus free song singing.
2. Repeat the study using a different and larger population to improve generalizability.
3. Repeat the study with a longer treatment period to see if additional treatment results in improved pitch accuracy.
4. Retest treatment subjects at various intervals after treatment to see if the treatment effects last.
5. Determine the reliability and validity of the singing stimuli.
6. Determine what the actual vocal pitch accuracy threshold is for the inaccurate singer by testing the electronic measurement with that of the ear of the music educator.
7. Repeat the study using singing stimuli in various keys to determine if vocal registration is a factor in singing inaccuracy.
8. Perform a longitudinal study to determine the magnitude of VPA fluctuation from grade to grade.
9. Conduct a study to determine how the Yuba Method compares with other vocal treatment methods. This would help to isolate the factor of additional instruction as the cause of singing improvement.

The implications of this study are that additional exercises such as those employed by the Yuba Method, can possibly help to correct inaccurate singing. This study demonstrates the value of working with inaccurate singers to improve vocal pitch accuracy.

## APPENDIX

### REFERENCES

- Aaron, J. C. (1991). The effects of vocal coordination instruction on the pitch accuracy, range, pitch discrimination, and tonal memory of inaccurate singers. *Dissertation Abstracts International*, 51(09), 2912A. (University Microfilms No. AAC 9103179)

- Apfelstadt, H. (1984). Effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. *Journal of Research in Music Education*, 32, 15-24.
- Bentley, A. (1968). *Monotones: A comparison with "normal singers" in terms of incidence and musical abilities*. Music Education Research Papers, No. 1. London: Novello.
- Brown, C. J. (1988). The effect of two assessment procedures on the range of children's singing voices. Unpublished master's thesis, Indiana University, Bloomington.
- Buckton, R. (1977). A comparison of the effects of vocal and instrumental instruction on the development of melodic and vocal abilities in young children. *Psychology of Music*, 5, 36-47.
- Campbell, M., & Greated, C. (1987). *The musician's guide to acoustics*. New York: Schirmer Books.
- Clayton, L. (1986). *An investigation of the effect of a simultaneous pitch stimulus on vocal pitch accuracy*. Unpublished master's thesis, Indiana University, Bloomington.
- Cleall, C. (1970). *Voice production in choral technique*. (Revised and enlarged). London: Novello.
- Collins, G. E. B. (2000). Breath-management instruction: Effects on the vocal accuracy and attitudes toward singing of sixth-grade choral students. (Unpublished Doctoral Dissertation, University of North Carolina at Greensboro).
- Cooper, N. A. (1995). Children's singing accuracy as a function of grade level, gender, and individual versus unison singing. *Journal of Research in Music Education*, 44(4), 222-231.
- DeYarman, R. M. (1972). An experimental analysis of the development of rhythmic and tonal capabilities of kindergarten and first grade children. In E. E. Gordon (Ed.), *Experimental research in the psychology of music*: 8 (pp. 1-44). Iowa City: University of Iowa Press.
- Eikum, R. L. (1963). Singer vs. non-singer: An educational research study of the Clarkston, Washington elementary school system concerning the students' ability to sing and factors affecting singing ability. Unpublished master's thesis, University of Idaho, Moscow.
- Flowers, P. J., & Dunne-Sousa, D. (1990). Pitch-pattern accuracy, tonality, and vocal range in preschool children's singing. *Journal of Research in Music Education*, 38, 102-114.
- Gackle, M. L. (1987). The effect of selected vocal techniques for breath management, resonance, and vowel unification on tone production in the junior high school female voice. Unpublished doctoral dissertation, University of Miami, Coral Gables.
- Goetze, M. (1985). Factors affecting accuracy in children's singing. *Dissertation Abstracts International*, 46, 2955A. (University Microfilms No. DAI8528488)
- Gordon, D. S. (1985). A survey of literature and practice in assisting the pitch-defective singer in the elementary school. *Pennsylvania Music Educators Association: Bulletin of Research in Music Education*, 16, 11-18.

- Gordon, E. E. (1984). *Learning sequences in music*. Chicago: GIA Publications, Inc.
- Gould, O. (1969). Developing specialized programs in singing in the elementary school. *Bulletin of the Council for Research in Music Education*, 17, 57-73.
- Gratton, M. (1992). The effect of three vocal models on uncertain singers' ability to match and discriminate pitches (Masters thesis, McGill University, Canada, 1989). *Masters Abstracts International*, 30 (4).
- Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary school children. *Journal of Research in Music Education*, 38, 225-232.
- Greenberg, M. (1970). Musical achievement and self-concept. *Journal of Research in Music Education*, 18(1), 57-64.
- Guerrini, S. C. (2002). The acquisition and assessment of the developing singing voice among elementary students. AAT 3040318.
- Haan, Charles, T. (1977). *Statistical methods in hydrology*. Iowa: Iowa State University Press.
- Jaffurs, S. E. (2000). The relationship between singing achievement and tonal music aptitude. Unpublished Masters Thesis, Michigan State University.
- Jarjisian, C. S. (1981). The effects of pentatonic and/or diatonic pitch pattern instruction on the rote-singing achievement of young children. *Dissertation Abstracts International*, 42, 2015A. (University Microfilms No. 8124581)
- Jones, M. (1971). A pilot study in the use of a vertically-arranged keyboard instrument with the uncertain singer. *Journal of Research in Music Education*, 19(2), 183-194.
- Jones, M. (1993). An assessment of audiation skills of accurate and inaccurate singers in grades 1, 2, and 3. *Update*, 11, 14-17.
- Joyner, D. R. (1969). The monotone problem. *Journal of Research in Music Education*, 17, 115-124.
- Kramer, S. J. (1985). The effects of two different music programs on third and fourth grade children's ability to match pitches vocally. Unpublished doctoral dissertation, Rutgers University, Rutgers.
- Levinowitz, L. M., Barnes, P., Guerrini, S., Clement, M., D'April, P., & Morey, M. J. (1998). Measuring singing voice development in the elementary general music classroom. *Journal of Research in Music Education*, 46, 35-47.
- Mathias, S. L. (1997). A teaching technique to aid the development of vocal accuracy in elementary school students. Unpublished Doctoral dissertation, Ohio State University.
- Miyamoto, K. A. (2003). *The report of a candidate for distinguished researcher medal*. University of Tokyo. Unpublished paper.
- Mizener, C. (1993). Attitudes of children toward singing and choir participation and assessed singing skill. *Journal of Research in Music Education*, 41, 233-245.

- Montgomery, T. (1988). A study of the associations between two means of vocal modeling by a male music teacher and third grade students' vocal accuracy in singing pitch patterns. Ann Arbor, Michigan: University Microfilms International 8822404.
- Moore, D. S., and McCabe, G. P. (2003). *Introduction to the practice of statistics, Fourth Edition*. New York: W. H. Freeman and Company.
- Music Educators National Conference (1994). *What every young American should know and be able to do in the arts: National standards for arts education*. Reston, VA: Music Educators National Conference.
- National Assessment of Educational Progress (1974). *The first national assessment of music performance* (Report 03-MU-01). Denver, CO: Education Commission of the States.
- Petzold, R. (1966). *Auditory perception of musical sounds by children in the first six grades*. Cooperative Research Project No. 1051, Madison: University of Wisconsin.
- Phillips, K. H. (1983). The effects of group breath control training on selected vocal measures related to the singing ability of elementary students in grades two, three, and four. Unpublished doctoral dissertation, Kent State University, Kent.
- Phillips, K. H. & Aitchison, R. E. (1995). Effects of psychomotor instruction on elementary general music students' singing performance. *Journal of Research in Music Education*, Volume 45, (2), 185-196.
- Phillips, K. H. (1996). *Teaching kids to sing*. New York, NY: Schirmer Books.
- Rainbow, E. L. and Hildegard C. Froehlich (1987). *Research in music education: An introduction to systematic inquiry*. New York: Schirmer Books.
- Richner, S. S. (1976). The effect of classroom and remedial methods of music instruction on the ability of inaccurate singers, in the third, fourth, and fifth grades, to reproduce pitches. *Dissertation Abstracts International*, 37, 1447A. (University Microfilms No. 76-19,898)
- Roberts, E., and Davies, A. (1975). Poor Pitch Singing: The response of "monotones" to a program of remedial training. *Journal of Research in Music Education*, 23, 227-239.
- Roberts, E., & Davies, A. D. M. (1976). A method of extending the vocal range of "monotone" schoolchildren. *Psychology of Music*, 4, 29-43.
- Rooks, J. M. (1987). The effect of range training on the vocal accuracy of restricted and complete range K-3 singers. Unpublished doctoral dissertation.
- Sims, W. L., Moore, R. S., and Kuhn, T. L. (1982). Effects of female and male vocal stimuli, tonal pattern length and age on vocal pitch-matching abilities of young children from England and the United States. *Psychology of Music*, Special Issue: *Proceedings of the IX International Seminar on Research in Music Education*, 104-108.
- Smale, M. J. (1987). An investigation of pitch accuracy of four- and five-year-old singers. *Dissertation Abstracts International*, 48, 2013A. (University Microfilms No. DA8723851)
- Small, A., and McCachern, F. L. (1983). The effect of male and female vocal modeling on pitch-matching accuracy of first-grade children. *Journal of Research in Music*

*Education*, 31(3), 227-233.

Snedecor, George W. (1956). *Statistical methods*, 5<sup>th</sup> edition. Iowa: The Iowa State College Press.

Welch, G. F., Howard, D. M., and Rush, C. (1989). Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, 17, 146- 157.

Wilson, D. S. (1970). A study of the child voice from six to twelve. *Dissertation Abstracts International*, 31, 5453.

Yarbrough, C., Bowers, J., & Benson, W. (1992). The effect of vibrato on the pitch-matching accuracy of certain and uncertain singers. *Journal of Research in Music Education*, 40, 30-28.

Yuba, Toru (1998). *Dream voice training: Muscles for singing* (with CD). Tokyo, JAPAN: Victor Entertainment, Inc.

Yuba, Toru (2000). Method for Curing Motor-Related Off-Key Singing: Is Tone Deafness Really Deafness? *The Japan Journal of Logopedics and Phoniatics*, 41, (4).

Yuba, Toru (2001). *How to Train the Muscles for Singing*. Okuyama, Japan. Unpublished paper.

Yuba, Toru (2002). Cause of Off-Key Singing and Its Curative Education. *Journal of Otolaryngology, Head and Neck Surgery*, Tokyoigakusha Inc., June.

Zwissler, R. N. (1972). An investigation of pitch discrimination skills of first grade children identified as accurate singers and those identified as inaccurate singers. *Dissertation Abstracts International*, 32, 4056A-4057A. (University Microfilms No. AAC 7202947)

---

[Current issue/Vol. 5](#) | [Vol. 4](#) | [Vol. 3](#) | [Vol. 2](#) | [Vol. 1](#) | [Purpose & Copyright](#) | [Submissions & Guidelines](#)  
| [Editorial Board](#) | [Contact Us](#)

ISSN 1532-8090

All material © 2007 *Research and Issues in Music Education* All rights reserved.