THE UNIVERSITY OF CALGARY

DEVELOPING MELODIC SINGING ACCURACY IN GRADE ONE STUDENTS

by

Elizabeth Fleming Boudreau

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF MUSIC

DEPARTMENT OF MUSIC

CALGARY, ALBERTA

OCTOBER, 1990

ISBN 0-315-66866-0

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommended to the Faculty of Graduate Studies for acceptance, a thesis entitled, "Developing Melodic Singing Accuracy in Grade One Students" submitted by Elizabeth Fleming Boudreau in partial fulfillment of the requirements for the degree of Master of Music.

Lois Choksy
Chairperson, Prof. Lois Choksy
Department of Music

Malcolm V Edwards.
Prof. Malcolm Edwards
Department of Music

Carol E. Harris
Prof. Carol Harris
Department of Music

A. L. Boberg
Dr. Alice Boberg
Department of Educational Policy
and Administrative Studies

Date: 19 October 1990

ABSTRACT

This study deals with the development of in-tune singing in young children and attempts to determine whether a particular sequence of exercises is effective in enabling Grade 1 children to improve their singing accuracy. Included is a review of the literature concerning the influence on in-tune singing of certain factors: melodic perception, home musical environment, gender, maturity, and instruction.

In each of two similar schools in Calgary, Alberta, two Grade 1 classes were selected and randomly assigned to either experimental or control instruction. For eight weeks, all four classes received 30-minute music lessons thrice-weekly, taught by the researcher. For 15 minutes of every lesson, all classes received instruction following the same Grade 1 music curriculum. For the other 15 minutes of every lesson, the experimental classes received instruction with Gould's Speech to Song Sequence, while the control classes sang additional songs that reinforced concepts in the curriculum. The experimental sequence consisted of speech inflection activities, "games" to distinguish the singing, speaking, and whispering voices, echo singing in a comfortable vocal range, exercises using the "oo" sound to extend the range, and echo singing patterns with neutral syllables then words. Experimental group students showed higher and lower sounds with arm motions.

Single note echo singing, melodic pattern echo singing, and song singing were measured with the Singing Achievement Measures(SAM), and melodic perception was measured with the Primary Measures of Music Audiation(PMMA). All students were pre- and posttested with these tests.

A two-way (instruction, gender) analysis of covariance with pretest covariate was used to determine the significance of the difference between group scores on the SAM and the PMMA after instruction.

Analysis of the scores yielded the following findings:

1. Grade 1 students who were taught with Gould's Speech to Song Sequence had significantly better single pitch echo singing, melodic pattern singing, and melodic perception after instruction than students who did not receive instruction with the experimental sequence. Grade 1 students' skills in song singing were not significantly improved by instruction with the Sequence.
2. The singing gains of experimental group students who were weak singers before instruction were significantly greater than the singing gains of control group students who were initially weak singers.
3. The singing gains of experimental group students with high melodic perception were significantly greater than the singing gains of control group students with high melodic perception. There was no difference in the singing gains of experimental and control group students with low melodic perception.
4. Before instruction, Grade 1 students with musical home environments had significantly better singing skills, but not melodic perception, than students with weak home musical environments.
5. Girls scored higher than boys on single pitch echo singing, melodic pattern singing and song singing after instruction, but the difference was not significant. Boys had slightly better melodic perception skills than girls, but the difference was not significant.

These components of the experimental program were considered by the researcher to be effective: instruction in vocal skills and concepts about pitch, the sequenced program, individual singing within a group and individualized programming, the exercises in the experimental sequence, the emphasis on careful listening, the opportunity for good singers to act as vocal models for their peers, and the combination of the experimental sequence and a program of traditional singing games, action songs, and listening songs.

ACKNOWLEDGEMENTS

I would like to thank Lois Choksy for her guidance in this thesis, and her dedication to music learning and music teaching.

The principals and Grade 1 teachers at St. Rita and St. Bede Schools, Richard Old, Louis LaPointe, Carole Baldwin, Jackie Parker, Lucy Assen and Barb Doll, willingly allowed for the participation of the children. I would like to thank them for this, and for their cooperation throughout the project.

I would like to express my thanks to several colleagues whose assistance has been invaluable: to Pat Hrynkiw, for her beautiful soprano voice, which provided a fine example for the children to follow on the singing test; to Darlene Ngo and Helen Beach, for their careful listening over many hours when evaluating children's singing on tapes; to Paul Domby, Francine Houlihan, Barney Munnalall, and Fiona Wilson for observing lessons; and to Mary Buckley, Luisa Izzo, Jean Auger-Crowe, Eila Peterson, and Jeannette Panagapka, whose friendship and support helped to carry me through the duration of the project.

A heartfelt thank-you is also extended to my sister Mary Patterson, for her encouragement, and the assistance of her family in the process of designing the singing test; and to my parents, for their love, and their assistance in my education.

A special thank-you is expressed to my husband Bernie, for his love, his constant support, and his wise advice; without his encouragement over several years and several thousand kilometers, this thesis may not have reached completion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Singing has been considered a basic form of musical expression throughout the centuries. Singing is an acquired skill, in a similar way that language is a learned skill. Without positive environmental influences, the singing capabilities of most children will not be developed to their potential, and although some children are fortunate to have musical encouragement at home, many others are not. The school environment can provide further opportunities for children to sing and experience the joy of music with others. However, there are many children who enjoy music and attempt to sing, but who have difficulty singing in tune. Music educators have felt a responsibility towards these inaccurate singers and have been attempting to create more effective ways to develop their students' ability to sing melodies accurately (Atterbury, 1984a; Bennett, 1986; Forcucci, 1975; Stene, 1969).

The Problem

Although numerous procedures for teaching children how to sing have been suggested by vocal educators, unless the success of these techniques is carefully evaluated, teachers' efforts may be inefficient, and children may not improve their singing. Researchers have investigated the effectiveness of some remedial and classroom programs in helping children to sing in tune. Some programs were found to be successful to a limited extent in controlled studies. In others, the approach contained such a wide variety of exercises that it was difficult to determine which techniques were effective. In several others, the experimental design was not a large priority, so that, although a wealth of information was provided, the results were not entirely dependable.

The Purpose

This study was designed to determine whether a specific sequence for teaching and learning was effective in improving the singing accuracy of Grade 1 students.

Question. Would there be any difference in the singing accuracy of two groups of Grade 1 students if one group's instruction included the Speech to Song Sequence of A. Oren Gould, and the other group received an identical curriculum except for the singing of additional songs instead of Gould's sequence?

Hypothesis. Grade 1 students who receive music instruction that includes Gould's Speech to Song Sequence of vocal skills and concepts will achieve significantly better singing accuracy and melodic perception skills than students who receive music instruction including the singing of songs without Gould's sequence.

Null Hypotheses.

1. There is no difference in single note echo singing scores between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.
2. There is no difference in melodic pattern echo singing scores between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.
3. There is no difference in scores for melodic accuracy when singing a song between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental

sequence.

4. There is no difference in melodic perception scores between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

Delimitations

The sample of subjects for this study was limited to children enrolled in four Grade 1 classes in northwest Calgary, Alberta.

Only accuracy of pitch and contour were measured in the singing tests used in this study. Vocal timbre, breathing, expressiveness and rhythmic precision were overlooked for the purposes of this study.

The measurement of single note echo singing was based on singing test scores on the Singing Achievement Measures Part I.

The measurement of melodic pattern echo singing was based on Singing Achievement Measures Part II test scores.

The measurement of melodic accuracy in song singing was based on Singing Achievement Measures Part III test scores.

The measurement of melodic perception was based on test scores on the Primary Measures of Music Audiation - Tonal. This test consists of Tonal and Rhythmic parts. Only the Tonal test was used in this study.

Definitions of Terms

In this study, several terms have particular meanings, as indicated.

"Singing accuracy" refers to precision in pitch and melodic contour when singing.

"Echo singing" is the imitation of one person's singing (usually the teacher's) by another (usually the student). This activity is also referred to as pitch matching or tone matching.

"Auditory perception" refers to the ability to perceive through the sense of hearing and includes rhythmic perception, harmonic perception, and melodic perception.

"Melodic perception" refers to the ability to perceive pitch; its components are "pitch discrimination" and "tonal memory".

"Pitch discrimination" refers to the ability to compare and recognize whether two pitches are the same or different.

"Tonal memory" refers to the ability to recall a melody after it is no longer audible.

"Pitch level" refers to the average absolute pitches in a song or melodic pattern.

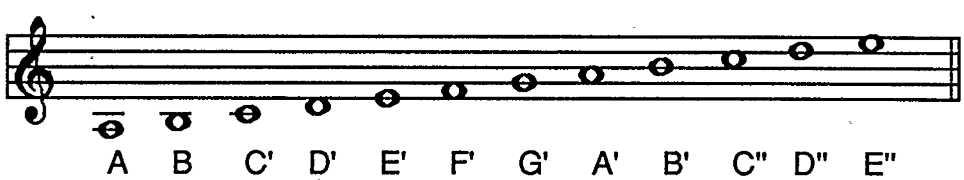
Each "experimental class" is an intact class, receiving experimental instruction.

Each "control class" is an intact class, receiving control instruction.

"Experimental group" refers to all students in both schools who received experimental instruction.

"Control group" refers to all students in both schools who received control instruction.

The notation of absolute pitches used in this study is:



CHAPTER II

REVIEW OF RELATED LITERATURE

Factors that Influence Singing Ability

Several factors may influence the present level of singing ability observed in children. In this review, the results of studies examining the role of melodic perception, home musical environment, maturation and gender will be discussed. Although these factors overlap, in the sense that maturation and previous musical experience probably influence melodic perception, they will be interpreted independently for present purposes. The effect of instruction will be discussed in the second section of this review.

Melodic Perception

Boardman (1964) observed that most of the Kindergarten, Grade 1 and Grade 2 children in her study could sing the general contour of a melody, although they were inaccurate in melodic detail. She suggested that they did have the necessary physical coordination to sing a melody correctly, but did not possess a fine enough sense of pitch perception to realize that small adjustments were required. Although Boardman's primary purpose was to investigate the influence of vocal training and maturation on vocal accuracy, she concluded that "improvement in perception, occurring as a result of additional musical experience plus maturation, is a factor in the development of vocal accuracy" (p. 84).

In 1968, Arnold Bentley reported the test results from the administration of his Measures of Musical Ability (MMA) to 349 "normal" singers and 53 "monotones" aged 9.7 years to 12.5 years, in twelve schools. This test measures a) pitch discrimination, by requiring that the second of a pair of tones is compared to the first by "up", "down", or "same", and b) tonal memory, by measuring the ability of a child to determine whether the second of two phrases is the same as the first and to show the location of the changed note by counting. Bentley reported that the difference between "monotones" and "normal" singers in both pitch discrimination and tonal memory was significant at the p = .10 level. These results concur with Boardman's observations.

In 1969, Joyner administered the Bentley pitch discrimination and tonal memory tests to 32 "monotones". Each had been assessed by their teachers as "someone who consistently fails to reproduce the tonal configuration of a melody in a recognizable manner" (p.115). He compared their scores with those of normal singers and found that the difference between the pitch discrimination of the two groups was significant. The raw scores indicated that the better singers also had better tonal memory skills than the "monotone" singers. He concluded, after his work with 11-year-old poor pitch singers, that there were three abilities necessary for vocal accuracy:

> 1) pitch discrimination, the ability to tell one pitch from another
> 2) tonal memory, the ability to recall a succession of pitches, and
> 3) a "vocal instrument capable of reproducing the succession of pitches in a melody" (p. 117).

Zwissler (1971) also compared the pitch discrimination of accurate and inaccurate singers. She found that the pitch discrimination skills of accurate singers in Grade 1 were significantly better than the pitch discrimination skills of inaccurate singers, over a range of three octaves. Pitch discrimination was measured as the ability to identify the second of a pair of tones as being "higher" or "lower" than the first; in this way her test was similar to Bentley's. Zwissler's results indicated a strong relationship between inaccurate singing and poor pitch discrimination skills. However, 10% of the inaccurate singers were found to have good pitch discrimination (scores of over 75% on the test); in these cases, some other factors were influencing singing accuracy. Zwissler recommended that the teaching of pitch discrimination skills occur at the primary level. Zwissler's results suggest that pitch discrimination training be included in any program designed to improve vocal accuracy.

Geringer (1983) compared the pitch matching of students with high, medium, and low pitch discrimination skills and found no significant difference in pitch matching skills among students in the three pitch discrimination groups. These results suggest that the level of pitch discrimination skill an individual has attained cannot be used as an indicator of level in singing skills. These findings were true for the 72 preschoolers and 72 Grade 4 students involved in the study. In addition, Geringer found the Spearman Rank Correlations between pitch discrimination and pitch matching to be quite low: $r = -.13$ for preschoolers and $r = .16$ for fourth graders.

Buckton, in his 1977 study of 41 six- to eight-year-olds, found a relatively low pretest correlation ($r = .11$) between vocal accuracy as measured by the Boardman phrase-singing test and pitch discrimination as measured by the Bentley MMA administered individually. A somewhat higher pretest correlation was found between vocal accuracy and tonal memory as measured by the three- and four-note tunes of the Seashore tonal memory test ($r = .23$). He also reported that seven-year-olds did not improve vocal accuracy scores, despite significant improvement in pitch discrimination.

M. Jones (1979) found that 16 of the 36 "uncertain singers" involved in her study scored average or above on the pitch discrimination pretest of the Bentley Measures of Musical Ability, but only four uncertain singers were average or above in tonal memory. In addition, most uncertain singers were able to determine whether two sounds were the same or different, but several subjects were unable to discriminate between high and low pitches. These results indicated that tonal memory is a more difficult skill and a much more important component of singing ability than single pitch discrimination. In the case of the four students with good tonal memory but poor singing, there must have been other reasons for their singing inaccuracies.

The results of Boardman, Bentley, Joyner, Zwissler, and Jones indicate a relationship between melodic perception and vocal accuracy. Direct correlations, such as those calculated by Geringer and Buckton, do not show a relationship between pitch discrimination and pitch matching. Jones' and Zwissler's findings also indicate that singing ability consists of other components in addition to melodic perception.

Home Musical Environment

Zimmerman, in her response to Gardner, Davidson, and McKernon (1979), stressed the need for a "rich and stimulating musical environment" from a very early age (p. 316). Moog's observations of preschool children's musical behaviors led him to assert that the singing and movement to music of two- to four-year olds was determined by innate ability and the influence of the environment (1976). The environment was viewed by Welch (1986b) as promoting or inhibiting the development of young children's musical behaviors. One very large component of a young child's environment is his or her home environment. The musical aspects of this environment comprise the home musical environment.

Kirkpatrick (1962) found a strong relationship between singing ability and home musical environment. The home musical environments of 116 five-year-old children were identified in three categories. The children's singing was tested, and the children were classified as singers, partial singers or non-singers. Specific situations found to relate to singing ability were: mothers who sang to and with their children, assistance in learning songs from parents and other adults, family participation in singing and playing instruments, and musical parents.

Shelton, in his 1965 study of 30 first grade children, also found a significant relationship between home musical environment and children's musical responses (singing, pitch discrimination, responses to rhythm, tempo, mood). Classroom teachers selected the most musical and the least musical children in their classes, and the researcher interviewed family members and church school teachers to determine what musical opportunities were available to the child. The factors that Shelton found to be most closely related to musical response were: frequent opportunities for the child to hear singing in the home and to sing with other members of the family, frequent opportunities to hear records at home, and the ability of the mother and father to sing and learn new songs.

The results of Moore's 1973 study showed that singing and rhythm responses of five-year-olds correlate positively with the following musical environment variables: musical instruments played in the home, parents and siblings who participate in musical activities, parental help with singing in tune and moving to music, and opportunities to hear various kinds of recorded music. Moore also noted that seven percent of the children who scored at or above the test mean were from non-musical homes, which indicated that other variables were influencing their singing and rhythm abilities.

Dibble's 1983 study examined the relationship between the home musical environment of five-year-olds and their ability to learn pitch discrimination skills. The pretest-posttest comparison indicated that although the students from both good and poor home musical environments improved equally in pitch discrimination, the posttest scores of the poor musical environment group were still lower than those from the good musical environment group.

Brand (1986) measured the home musical environment of Grade 2 students using a parent self-reporting questionnaire. The subjects' tonal and rhythmic perception were evaluated with the Primary Measures of Music Audiation (Gordon, 1979a), and their musical achievement (musical knowledge, skill, music reading, and musical initiative) was determined by the school music teachers and recorded on an assessment form. The results of the three measures were correlated. Results indicated a strong relationship between home musical environment scores and musical achievement of second-grade children. There were no significant correlations between factors of home musical environment and tonal or rhythmic perception scores. The strongest relationship found was between musical achievement and two aspects of home musical environment: overall parental attitudes toward music and parental involvement musically with the child.

The findings of Kirkpatrick, Shelton, Moore, Dibble, and Brand concur that there is a positive relationship between children's singing and other musical behaviors, and certain home environmental factors. In particular, there seems to be a connection between children's singing abilities and the frequent singing of parents and other family members.

Maturity

Singing and Pitch discrimination Skills at Different Ages.

Maturation and experience both have roles in development. Because maturation is difficult to measure, age is used as an indicator of level of maturity.

Bentley found a "small but steady increase" (1966, p.108) in mean scores with increasing age (7 years to 14 years) on the pitch discrimination and tonal memory tests.

Petzold (1963) described his 45-item test as a test of auditory perception of short tonal configurations. The children responded by attempting to vocally echo each pattern; therefore, the test essentially measured singing skill, even though a component of this skill is auditory perception. In the sample of 606 children in Grades 1 through 6 in Madison, Wisconsin, the results showed a higher test mean for each successively higher grade. T-tests showed no significant difference between test scores in Grade 1 and 2, although significant differences were found between Grades 1 and 3. Significant differences were also found between Grades 5 and 6, 4 and 6, 1 and 4, 2 and 5, and 3 and 6, suggesting that there are three populations of singers of elementary school age: Grades 1 and 2, Grades 3 through 5, and Grade 6. Unfortunately, Petzold's results are not entirely reliable because of his use of several t-tests where an initial analysis of variance would have been appropriate; this increased his likelihood of finding statistically significant differences where none would, in fact, exist.

Geringer (1983) tested the pitch discrimination and vocal pitch matching skills of 72 preschoolers and 72 Grade 4 students. The results showed a significant difference between the two groups of older and younger children, on pitch discrimination and vocal pitch matching.

Gould (1968a) surveyed music teachers across North America to find out what percentage of the children they taught had singing difficulties. Averages of the 602 teachers' estimates were as follows: First Grade - 34.6%, Second Grade - 24.2%, Third Grade - 17.8%, Fourth Grade - 12.9%, Fifth Grade - 11.8%, Sixth Grade - 11.0%, all grades - 18.7%.

Davies and Roberts (1975) contacted school music teachers in Chester, England, and requested that they classify each of their students as either a "normal singer" or a "poor pitch singer". The results of 10,646 five- to eleven-year-olds showed that poor pitch singing tended to decrease with age. The percentages of poor pitch singing in each age group, beginning with the five-year-olds were: 44%, 38%, 27%, 22%, 23%, 18% and 18%.

All these results indicate that children generally improve both in their melodic perception and in their singing ability as they mature. To what extent their improvement is influenced by their inherent aptitude and by the previous musical activities they participated in, cannot be determined from the given data.

Age to Begin Vocal and Melodic Perception Instruction. Jersild and Bienstock (1934) found that the singing of 23 children (three to eight years old) improved at least 30% after two months of daily training. The researchers reported that when the children were tested two years later, they had still retained the benefit of earlier training. However, because only twelve of the original 23 could be located for posttesting and there were poor experimental controls, Jersild and Bienstock recognized that a definite conclusion should not be made from the results.

Smith (1963) found that vocal training in groups of 20, improved the singing of three- and four-year olds. At the end of 32 weeks of daily instruction, the experimental group had a higher proportion of tuneful singers than the proportion of tuneful singers in the control group.

Boardman (1964) found that children in Kindergarten, Grade 1 or Grade 2 who had not had vocal training in preschool, were at a similar level of singing accuracy as the children in Smith's study who had received preschool training. She theorized that preschool vocal training may accelerate the developmental process, but does not affect the end-product (p. 80). As stated earlier in this review, Boardman believed that perception was important in the development of vocal accuracy.

Gould (1968a) concluded that early childhood is probably the most effective time for teachers to help students with singing problems. Similarly, Joyner (1969) found that many potential monotones could be helped through "early training in the easy command of the vocal instrument" (p. 124). He recommended that training begin early to prevent poor vocal habits from continuing.

Petzold (1969) reported that there are indications that the most significant changes in auditory perception (measured in vocal response) occur between Grades 1 and 2. "Unless greater attention is paid to the development of aural understandings when the child is in the first grade, this will seriously inhibit his subsequent musical development" (p. 85).

Buckton (1977) suggested that when children are introduced to auditory skill training between the ages of five and eight years, the level they will eventually attain will be higher than if their training had begun later.

Results concerning exactly when to begin vocal instruction are inconclusive, possibly because the tests used do not measure exactly the same characteristics. This demonstrates the need for adequately standardized tests of singing skill and melodic perception for children, especially those under eight years of age.

Although there is some disagreement, all of the above educators seem to think that training should begin at a young age - preschool or early elementary school. Jersild and Bienstock, Smith, Gould, and Joyner encouraged vocal training; Boardman, Petzold, and Buckton emphasized the importance of training melodic perception skills.

<u>Gender</u>

In Jersild and Bienstock's study (1931) of 48 three-year-olds, boys were found to have a wider vocal range than girls. They report that because of the small number of subjects, the apparent higher skill level of boys may be due to chance. In 1934 they tested the vocal range of 407 children, 2 to 10 years old. Their findings indicated that at all ages except two years, girls were able to sing more notes than boys although the difference was not statistically significant.

The auditory perception of 660 students was measured with Petzold's test (1963) as previously described (that is, with a singing response), and girls were found to have higher mean scores than boys at each grade level except Grade 6. However, t-tests showed no significant difference between boys and girls on any of the subtests, including the phrase singing test.

In Gould's project (1968a), the ratio of boys to girls needing help with their singing increased at each higher grade level. At Grades 1 and 2, the number of boys and girls needing help was almost equal, but by the Grade 6 level, the ratio was 12 to 1.

Moore (1973) tested 101 five-year-olds and found that girls performed significantly better than boys on the vocal range and pitch accuracy subtests.

Davies and Roberts (1975) surveyed approximately 5476 boys and 5170 girls aged 5 to 11 years and found the overall incidence of poor pitch singing to be 36% for boys and 18% for girls. This difference was found to be highly significant ($p = .001$). In addition, the incidence of poor pitch singing in girls was lower than in boys within every age group. The percent of inaccurate singing by sex and age was as follows:

	5yrs	6yrs	7yrs	8yrs	9yrs	10yrs	11yrs
Boys	51%	47%	37%	32%	31%	24%	26%
Girls	37%	30%	18%	11%	13%	12%	10%

Davies and Roberts' results showed that the proportion of boys with singing problems to girls with singing problems increased with age; boys seemed to be less likely to overcome singing problems as they got older. School music teachers classified poor pitch singers into five descriptive categories of singing problems:

- can sing correctly part, but not all of the melodic line,
- can sing the melodic line correctly, but at a lower pitch,
- can sing the melodic line correctly, but at a higher pitch,
- does not follow the melodic line at all, pitch very erratic, and
- "monotones", always completely untuneful with little variation in pitch.

There were more boys than girls in every category except the category where the melodic line was sung too high. The singing of 10% of the boys and 3% of the girls was classified in the lowest category and this difference was found to be significant at the $p = .001$ level. According to Davies and Roberts' survey, more boys have serious singing inaccuracies.

In her analysis of Kindergarten, Grade 1 and Grade 3 children's ability to echo sing melodic patterns with certain features, Goetze (1985) found that girls sang more accurately than boys. Students sang more accurately individually than in a group, and the difference between boys' individual and group responses was greater than that of the girls.

The results of Jersild and Bienstock, Gould, Moore, and Davies and Roberts indicate that, overall, girls have better singing skills than boys, although boys and girls appear to be more similar in their singing abilities at a young age. Since it seems unlikely that boys and girls have inherent differences in singing ability, some of the factors that promote or inhibit singing accuracy must be somewhat different for girls and boys.

Bentley (1966) tested 590 boys and 566 girls aged eight to twelve and found no significant difference in pitch discrimination and tonal memory between boys and girls, across all age groups.

Apfelstadt (1984) found no significant difference between Kindergarten boys and girls on the pre- and posttests of pitch discrimination, pattern singing and rote singing in the 23 boys and 38 girls she tested.

The results from Apfelstadt's singing tests do not agree with the studies above that dealt with singing skills, but the results of Bentley's and Apfelstadt's melodic perception tests do concur. Although boys seem to have adequate melodic perception compared to girls, boys have not acquired singing skills as well as girls. These findings suggest that effective singing programs for boys should include somewhat different components than programs for girls.

II Programs for Training Singing Accuracy

Many training programs have been undertaken in an effort to improve the melodic accuracy of children's singing. These programs have employed a variety of techniques.

Programs that emphasized the teaching of melodic perception

Several researchers have employed techniques to improve students' melodic perception and pitch concepts in a effort to improve students' singing accuracy.

The problem of poor pitch discrimination in the uncertain singer was addressed by B. Jones (1981). Bentley's MMA was used as a pre- and posttest of pitch discrimination. For six weeks Grade 2 students received instruction daily for twenty minutes. The purpose of the experimental program was to develop pitch awareness in each child, through the use of body motions which spatially reinforced pitch concepts while singing. The control class did not kinesthetically or spatially reinforce pitch. Both classes received instruction in speech chants, echoing environmental sounds, tone matching games, scale songs, and question and answer games based on the minor third interval. Both groups improved considerably as a result of instruction, and although the experimental group improved more, the posttest difference was not statistically significant. It appears that the students' pitch discrimination improved as a result of the singing activities participated in by both groups, in addition to the influence of spatial reinforcement in the experimental group.

Techniques to spatially, visually and kinesthetically reinforce pitch concepts were also employed in Steeves' 1984 research. Specifically, the Curwen-Kodály handsigns were used by Grade 4 experimental group students to reinforce concepts about interval direction and size. The handsign group's pitch discrimination posttest scores were significantly higher than the no handsign group's scores, as measured by the pitch discrimination test in Bentley's MMA. The results also indicated that the handsign group's tonal memory improved more than that of the no handsign group, but not to a significant extent. It would be interesting to know to what extent students' singing improved; unfortunately no singing test was given.

Jordan-DeCarbo (1982) taught two groups of kindergarten children for eleven weeks, three times a week, and measured their pitch discrimination with the Primary Measures of Music Audiation (Gordon, 1979a), and their singing with an investigator-designed singing test. The same/different discrimination group was involved in decision-making about whether melodic patterns were the same or different. The focus was on pitch concept development in an aural and verbal approach. The group receiving no same/different discrimination techniques echoed the same tonal patterns, but there was no discussion of similarity or difference in patterns. Both groups imitated the same eight tonal patterns sung by the teacher, but no vocal techniques to improve singing accuracy were employed. The Gordon test (PMMA) requires children to indicate whether the second of two patterns heard is the same or different from the first. Because all eight patterns taught were test items on both the PMMA and the singing test, one would expect to see an improvement in the students' test scores. In fact, both groups improved their singing and pitch discrimination scores, but not to a statistically significant extent. Possibly, the echo singing activities helped the pitch discrimination and singing of students in both groups. The results seem to indicate that verbal discussions alone are not sufficient to influence the singing and pitch discrimination of kindergarten students.

Apfelstadt's research (1983) was concerned with the effects of melodic perception instruction on the pitch discrimination and singing of kindergarten students. For 11 weeks she taught two groups, (labelled E1 and E2) in twice-weekly, 30-minute classes, and another teacher taught one additional class (Control) for the same length of time. Both experimental classes echo-sang many tonal patterns and songs, but the E1 group reinforced vocal patterns with hand levels, body movements and bell-playing on stepbells. Verbal reinforcements and visual icons were also used. The E2 group used no visual, kinesthetic or verbal reinforcements of pitch direction. Instead, rhythmic concepts were emphasized with icons showing duration and by clapping and playing rhythm patterns on non-pitched percussion instruments. The Control group instruction was participatory and activity-oriented and was not conceptually based.

The class (E1) that received melodic perception instruction did not have significantly higher posttest scores than the class (E2) that was taught the same program without the reinforcement of melodic direction and contour, on either the Boardman pattern-singing test (1964), the rote song test, or the PMMA. This indicates that the techniques used were not found to be effective in improving children's pitch discrimination and singing. Possibly the instruction in the two groups was too similar to effect a difference in singing accuracy. Apfelstadt suggested that the focus on melodic rhythm in the E2 group may have aided singing accuracy, which would have brought the E2 scores on the singing tests closer to the E1 scores. There was a significant difference in the posttest results of the E1 class and the Control class on the Boardman test, and between the E2 class and the Control class on the same test. However, because the experimental classes and the control class were taught by two different teachers in two different schools, the difference in results could be due to different teaching styles.

In 1933, Wolner and Pyle reported their efforts to improve pitch discrimination and singing in seven Grade 5, 6 and 7 students. These students could not distinguish differences as large as octaves, fifths, thirds, and they could not sing. For three months they received individual instruction twenty minutes daily. The meaning of high and low pitch was emphasized, analytical thinking about pitch was encouraged, and training in echo singing and pitch discrimination was employed. Students listened carefully to the piano, and sang single pitches, then stepwise sequences, then increasingly larger intervals to the syllable "la". After students were able to distinguish larger intervals, they were required to identify as higher and lower, two tones less than a semi-tone apart played on whipple forks. Individual differences in students were responded to, and great patience was demanded of the teachers. After three months, all pupils could discriminate octaves, fifths, thirds, tones, semi-tones over a wide range. All substantially improved their singing; exercises and songs without words were the most successfully sung. Improvements were observed after training in careful listening, echo singing, and pitch discrimination/pitch concepts. The amount of time and intensity of practice received in daily, individual training was probably a factor in their marked improvement.

Thirty-six "uncertain singers" from Grades 2, 3 and 4 were taught individually by M. Jones (1979) in daily, fifteen-minute sessions for twelve days. Three different teaching procedures were compared: vertical keyboard, horizontal keyboard, and no keyboard. A sequential procedure for singing improvement was followed for both keyboard groups. For students who received vertical keyboard instruction, the keyboard was arranged vertically and used to reinforce pitch concepts visually, spatially and kinesthetically. In the horizontal keyboard group, no visual or physical representation of "high-low" was given when playing the keyboard. In both keyboard groups, all melodic patterns were played on the keyboard before singing. In the no keyboard group, the same basic concepts and songs were taught, but the child echoed the investigator's vocal presentation of the patterns. Careful listening and accuracy in singing were encouraged in all students. Songs of a limited range containing the descending minor third or patterns descending stepwise to the tonic were used initially, progressing to songs with a wider range. Students were tested with an aural-vocal skills test, and the analysis of covariance with pretest covariate showed a significant difference between the posttest scores of the vertical-keyboard group, and the other two groups at the $p = .10$ level. The ANOVA on gain scores showed that the gain of the vertical-keyboard group was significantly different ($p = .05$) from the other groups. Jones observed that the vertical keyboard procedure was most effective in dealing with problems related to pitch direction, lack of attention and a low speaking voice.

Welch (1984) used an oscilloscope screen with seven-year-olds to show their vocal responses as a visual trace which provided a visual feedback of their sung pitch. Students received "knowledge of results" (1985, p. 6) on the screen by comparing a coloured target indicating the stimulus pitch with a mark representing their singing. This provided a rating of their singing, and visually reinforced their pitch concepts about the higher and lower sounds they sang and heard. Subjects receiving both visual feedback and knowledge of results had significantly higher singing test scores than students receiving no feedback or visual feedback alone. Without knowledge of results, students did not have accurate information about the sounds they were producing. Students had greater pitch matching success when a variety of pitches were presented as stimulus pitches ("variability of practice" 1985, p. 15), than when the same stimulus pitch was repeated. Although the use of the oscilloscope demands that Welch's procedure be an individual one, it may be possible to apply the theory about knowledge of results to a classroom setting.

Of the above studies, only Welch, M. Jones, and Wolner and Pyle were able to show an improvement in students' singing. The vertical keyboard was used by M. Jones to reinforce high/low sounds and as a visual and physical representation of the spatial relationships of pitch. Steeves reinforced specific pitches using a spatial and kinesthetic technique and students' pitch discrimination was improved. Apfelstadt and B. Jones also used visual, spatial and kinesthetic representation of pitch, but their programs were not as effective. Possibly the spatial reinforcement used did not represent pitches precisely enough, or possibly the experimental group's instruction was not sufficiently different from the control group's instruction. In addition, M. Jones' vertical keyboard program included the reinforcement of correct responses with a light, was highly participatory, and was probably quite motivating for the child. Jordan-DeCarbo's experimental group reinforced a very basic pitch concept, but the kindergarten children did not respond to the solely verbal reinforcement. The experimental and control programs may have been too similar to produce significantly different singing results. Both Wolner and Pyle's program and

M. Jones' program were sequenced, beginning with pitch matching and gradually extending the range and length of pattern. Although Wolner and Pyle focused on pitch discrimination and high/low training, their approach was vocal. Welch's conditioning program with visual reinforcement of pitch concepts and knowledge of results successfully improved children's pitch matching. Whether the effectiveness of knowledge of results and visual feedback can be extrapolated from pitch matching in a highly controlled individual setting with refined apparatus, to singing longer patterns and songs in a group setting, is a matter for future research.

Programs that focused on vocal training

Other researchers have emphasized the use of vocal techniques to teach singing accuracy.

Smith (1963) taught three- and four-year-old children in two groups, attempting to demonstrate that young children could learn to sing in preschool settings. In daily 15-minute sessions, each experimental group sang folk songs containing repetition of words or melody in different phrases. Songs were sung as long as the group maintained interest. After 32 weeks, the experimental groups, and control groups of similar children who did not receive instruction, were tested on their ability to echo short song phrases. Unfortunately, the rating scale was imprecise as it required that each student's singing be categorized as either tuneful or untuneful. After training, both experimental groups had a higher proportion of tuneful singers than their corresponding control groups.

Richner's research (1976) attempted to determine the success of a remedial singing program with Grade 3, 4 and 5 inaccurate singers. Four different training programs (Treatment) were used, each one at a different school. All groups at each grade received music instruction in two 25-minute lessons per week. In Treatment I the whole class received music lessons taught by the classroom teacher. Treatment II was music instruction taught to the whole class by an elementary music specialist. Treatment III consisted of singing songs in small groups of ten mostly inaccurate singers, taught by an elementary music specialist. Treatment IV consisted of remedial voice training in small groups of ten mostly inaccurate singers, taught by an elementary music specialist. All the inaccurate singers in each group were pre- and posttested with a singing test designed by the researcher. Richner's training program contained the following components:

- individual exercises,
- single note pitch matching, starting on middle C,
- echo D" - B', "yu-hu", "cuckoo",
- fire sirens in head voice, sustaining final note,
- child who is not reaching higher pitches, is asked to sing louder,
- D"C"B♭A'G', slur on syllable "oo" to sound like wind on a scary night (to bring head voice into chest voice register),
- sustained sound "lu", held as long as possible, echoing "lu" on short melodic patterns made up entirely of steps,
- "Hoo" owl echo on D". "Listen" first, then sing higher and lower by step, gradually working higher,
- sing vowels to notes of song "ay" or "oo", then change to words,
- echo with "ha", individual pitches played on piano getting progressively quicker.

There was no emphasis on melodic concepts; in fact, the students may have had some confusion about higher/lower pitch concepts because they were asked to sing louder when they had difficulty singing higher. The teacher provided positive feedback, calling attention to correct responses only, changing the exercise if responses were incorrect. Students were encouraged to take pride in their progress and to offer cooperative support for each other.

Results showed that at the Grade 5 level, inaccurate singers receiving remedial treatment (T. IV) improved significantly on their singing test scores over inaccurate singers in every other treatment group. Grade 4 students showed no significant differences between any treatment group on the posttest. In Grade 3, inaccurate singers receiving remedial small group instruction (T. IV), or small group instruction in song singing (T. III), improved significantly over inaccurate singers in regular classroom (T. I), but not over inaccurate singers in the regular classroom taught by a music specialist (T. II). The results across grades are not consistent, nor do they suggest a trend due to age or maturity, because the Grade 4 results were quite different from one grade higher or one grade lower. Due to some weaknesses in the design of the study, where each treatment was assigned to a different school and the cell sizes for the ANCOVA were quite small, the results are not entirely reliable. It can be cautiously concluded that the remedial vocal training in small groups may have had some positive influence in improving singing accuracy.

In a well-designed research project, Phillips (1985) evaluated the influence of breath-control training on the singing ability of students. The training included physical-conditioning warm-ups, breathing and breath-management exercises, and tone production vocalises and exercises. The experimental group scores on vocal pitch range, vocal intensity, and vocal accuracy were significantly higher than the control group scores. Group breath-control training was very effective in improving the singing ability of Grade 2, 3 and 4 students.

Forcucci (1975) classifed singing abilities in four categories. The independent singers are those who are able to sing accurately alone without accompaniment. The dependent singer can sing satisfactorily in a group, but has difficulty in more challenging situations. Forcucci suggests that with more experience, motivation and greater individual effort, dependent singers can become independent. Uncertain singers sing out-of-tune to varying degrees in all situations. Restricted-range singers have difficulty in producing more than one or two different pitches.

Forcucci successfully used several techniques in weekly, individual training with 186 Grade 3 to 6 students. Training began with pitch matching using the syllable "wa" in a comfortable range. Students echoed short patterns in the five-note toneset from tonic to dominant and proceeded to simple songs in this toneset. From the beginning, pitch concepts were reinforced aurally and visually at the piano keyboard. Students were asked to discriminate whether their singing was the same or different from the piano notes, and higher or lower adjustments were suggested. Initially, 90% of the students taught were uncertain singers, 8% were dependent singers, and the others were resticted-range singers. After 20 weeks, all except nine participants were independent singers.

Joyner (1969) used a very careful classification process to evaluate 11- and 12-year-old "monotone" singers and describe their vocal qualities and difficulties. Twenty minutes of individual training was given to eight problem singers, four times a week for 15 weeks. Joyner had the students attempt to match single pitches. Then, vocal production exercises designed to improve vocal resonance and breath control were employed. The singers then used a glissando to gradually expand the singing range to five notes. The "ah" vowel was used and the singers learned to echo short song phrases with a normal legato. Although Joyner thought that vocal control is not the only skill necessary to be able to sing in tune, he thought that training the voice is more effective than training in pitch discrimination and tonal memory alone. He also stressed the importance of beginning singing training in the primary grades.

One of the most comprehensive research projects concerning children's singing problems was completed in 1968 by A. Oren Gould. Gould is frequently cited as a proponent of the Speech to Song approach, which involves the gradual acquisition of singing skills after an initial introduction using speech inflection exercises (Atterbury, 1984b; Goetze, Cooper & Brown, 1990; Roberts & Davies, 1975; Rosborough, 1972; Welch, 1979a). His project, conducted over a three-year period, was in three phases.

In Phase I, surveys about children's singing problems were returned by 602 music teachers at the elementary and college level across North America. He attempted to determine the specific singing problems that out-of-tune singers had, and the techniques that teachers had found to be successful in improving the children's singing.

Opinions about singing problems lead to the following classifications:

1) the "too low singer",

2) the "too high singer",

3) the "one note singer",

4) the singer whose problems were a combination of all three, and

5) the psychologically inhibited singer.

Ideas about the causes of singing problems, listed in order of frequency reported by teachers were:

1) inattention to pitch and failure to notice pitch changes,

2) psychological inhibitions toward singing created by environmental impacts,

3) inability to coordinate the vocal mechanism with pitches heard,

4) low speaking voice,

5) lack of interest in singing attributed to causes such as poor teachers and materials,

6) lack of practice in singing, and

7) lack of exposure to music at home.

Remedial techniques suggested by teachers surveyed were:

1) tone matching drills,

2) use of speech devices,

3) use of bodily movements,

4) use of song pattern devices,

5) use of mechanical devices such as piano, bells, recorders, etc., and

6) miscellaneous activities including humming; whistling, siren and sound effects, listening experiences, group participation, placing a beginning singer near a strong singer, atmosphere-encouragement, imitation-echo, pitching songs within speaking range of the child, and individual attention.

In response to the question about when to help problem singers, 34% of the respondents said the best time was in Grades 1, 2 and 3, and 11% said Grades 4, 5 and 6. Nine percent felt that Kindergarten was the best time and the remainder thought that some children need help all through school.

Gould observed teachers and pupils in several schools from preschool to Grade 6 across the United States. As a result of the observations, conferences, and survey suggestions made in Phase I, a collection of experimental remedial activities was compiled.

The singing test to be used in the project, called the Gould Speech and Song Response Test, was developed and reached its final form by the beginning of Phase II. The child was required to echo a taped speaking or singing voice in the first five sections: speech with inflection, a short melodic phrase, a 6-note phrase with "loo", the same phrase sung with words, and individual pitches sung to "loo". The test of melodic perception required the child to tell whether the second of two notes played on the piano was higher, lower, or the same. The student was also asked to sing two songs of his or her choice. A rating scale was used to score the responses, and judges evaluated the song tapes. Some revisions in the content of the Gould Speech and Song Response Test were made between the second and third phases, and the reliability coefficients for the singing test, computed for each grade at the end of Phase II, were $r = .92$ or better.

The second phase of Gould's research involved a quasi-experimental pilot study with Grades 1, 2 and 3 at one school to test the techniques that were suggested most frequently by teachers during Phase I. In the experimental classes, experimental techniques and speech to song response materials were used for 10 minutes twice weekly as part of the regular music class. The control groups received the same program except for the experimental remedial procedures. The experimental program at this stage was based on the observations and survey results from Phase I, and appears to have been quite flexible. Adjustments were made to the program by the teachers involved and these additional techniques were included in the experimental sequences used in Phase III. The use of the syllable "oo" was added as a means of expanding the upper range of the "too low" singer. Singing exercises using more open vowels were also added to open up the voices of children who sounded pinched or breathy and to provide a smoother transition from the "oo" syllable to melodies with words.

All students were pretested with the Gould singing test, and after 18 weeks of instruction, they were posttested. A gain score, which is the difference between pretest and posttest scores, was calculated for each student.

A comparison of pre- and posttest means revealed that the experimental classes always scored higher than the control classes on the posttest. The difference between the experimental mean gain and the control mean gain was found to be significant at the $p = .05$ level for Grades 2 and 3, and at the $p = .001$ level for Grade 1.

The Grade 1 experimental class had been selected for their extremely low pretest scores and when this occurs, higher posttest scores are to be expected, due to the effect of regression. However, the experimental class improved to such an extent that their posttest scores surpassed the control class posttest scores, which suggests that the treatment, not statistical regression, was a strong influence. This class made larger gains than the Grade 2 and 3 experimental classes, and the project staff believed that there were two reasons for this: the use of more effective materials and techniques in Grade 1, and the receptiveness of Grade 1 children to the exercises.

Although the experimental classes improved more than the control classes, the inconsistencies in the experimental program make it difficult to determine which components of the program were effective. The control classes also improved considerably, which Gould suggested could be due to maturation factors or to an increased interest on the part of the control group teachers in improving children's singing.

Unfortunately, the project report does not outline the experimental program used in this phase of the project. The results of the pilot study showed that there was sufficient value in the experimental techniques to proceed to a larger-scale experiment.

In addition to the pilot study research at Phase II, individual case studies without control groups were conducted in seven other schools. The teachers were instructed to use techniques they had found to be successful and to create new exercises as necessary. A wide variety of procedures were used, and several were incorporated into the experimental teaching sequence in the next phase. Some of the successful activities and materials were:

- musical "conversations" moving from low to high to low pitches,
- speech activities using songs and poems to develop vocal control,
- siren imitations to get the voice into a higher range,
- "meow" exercises to find head voice,
- body motions to show higher and lower sounds,
- listening and identifying which of two tones is higher,
- songs with scalewise patterns and repeating phrases, and
- songs with a long skip to a high tone.

Tape recordings and teachers' reports documented increased singing accuracy and better tone quality in the children's singing. The case studies resulted in a collection of procedures found to be helpful in improving children's singing.

During Phase III, quasi-experimental studies were conducted with Grades One through Six in six schools, as well as additional case studies. There were control and experimental classes at each grade level, and the study lasted for 18 weeks. Students were pre- and posttested with the Gould test. The experimental group teachers were instructed to use the teaching techniques found to be helpful in the Phase II research and to adapt the program to song material they had chosen. They were also encouraged to add their own techniques and revise those suggested. The control classes used the same song materials without the experimental sequence. Each experimental class received instruction with the experimental sequence for 10 minutes in music class twice a week, and some problem singers received extra assistance in individual or small group practice. Results showed that the experimental group improved significantly more than the control group, at the $p = .05$ level.

In Phase III, the experimental program appears to have been defined more clearly for the teachers, although they were still encouraged to try new techniques they thought might be helpful. These techniques were incorporated into the program, and the sequence was finalized at the conclusion of the project. It was basically a sequence of skills and concepts designed to be developed together to promote in-tune singing. Gould recorded the sequence in a manual Finding and Learning to Use the Singing Voice (l968b). See Appendix A for Gould's Speech to Song Sequence. Phase III case studies using experimental techniques provided information about additional procedures found to be effective. Many activities were similar to those used in Phase II, and all the case studies noted substantial improvement in singing accuracy.

Although skepticism has been expressed concerning Gould's use of gain scores in his data analysis (Phillips, 1984), there are still some writers on the subject who do not dismiss gain score analysis (Best, 1986; Burroughs, 1971; Cook and Campbell, 1979). In any case, the significance levels reached in the Phase III experiment were sufficiently high (p = .001 for Grades 1, 2 and 5, p = .01 for Grades 3A and 6) to speculate that significant differences might also have been found if gain score analysis had not been used. Unfortunately the report does not specify whether an analysis of variance or t-tests were used for significance testing.

However, other uncontrolled factors in the experiment raise questions about the dependability of the results. No attempt was made to control the amount of instructional time that problem singers received, as some experimental group students received additional help. The extra practice in and of itself may have caused the improvement. In addition, it is not clear whether experimental and control classes were taught by the same teacher. If they were not, then teaching style rather than teaching procedures may have caused singing improvement.

The experimental group teachers used different techniques and all the experimental groups were not taught the same singing skills using the same teaching procedures. Although the experimental groups showed significant gain over the control groups, it is difficult to determine exactly which techniques were effective in improving the children's singing. The report mentions that control group teachers may have used some of the experimental techniques, and it appears that they were not cautioned against this.

Gould (1968a) was aware of the uncontrolled influences in the project. His purpose seems to have been to find ways to improve the singing of as many children as possible, rather than to find out precisely which techniques were effective. However, the research generated considerable qualitative data concerning the development of children's singing skills, and the results definitely show that children can improve their singing with practice. The results and the observations of the project staff led Gould to conclude that the sequential formation of concepts about pitch relationships and about the singing voice, and the development of vocal skills, is vital in the process of learning to sing (p. 49). Gould believed that children with singing difficulties need to understand the concepts of singing, after which their vocabulary of aural, mental, and vocal skills must be developed.

In an effort to build upon Gould's work, Roberts and Davies (1975, 1976) designed a controlled experimental study in an attempt to improve the singing of children with substantial pitch inaccuracies, through individual and small-group remedial training. The 90 six- to eight-year-olds involved in the project were all classified as "monotones" according to Davies and Roberts' 1975 categorization described previously in this review. They were chosen randomly from five schools and were randomly assigned to one of three types of training: Remedial, Traditional, or Control. The Control group received the usual singing lessons in their school, and the only thing that singled them out from the other students in their classes was the pre- and posttesting they received. The Traditional students received 30 minutes of small group and individual instruction twice-weekly, in addition to the regular music lessons at their school. They sang several songs with piano, guitar and percussion accompaniment. The Remedial students received their regular music lessons at school as well as 15 minutes of individual training and 15 minutes of group instruction twice-weekly. Ten music education students taught both the remedial and the traditional training programs. The remedial program consisted of a structured sequence of activities beginning with speaking and chanting, then tone-matching drills within the speaking range, developed from Gould's work:

1. Initially, speech devices were used to extend the range of spoken pitch. Short phrases and nursery rhymes were imitated with subdued or exaggerated inflection.
2. The transition from speech to singing was begun with single pitches. Each child attempted to sustain "lah" on one note, repeat the same note, then play it on the piano. From this "personal note" the child's range was slowly expanded. Semi-tones higher and lower than the personal note were attempted, repeating the pitch on different rhythm patterns and on the child's name. In the group setting, children took turns playing "their" note on the chime bars, then singing it. The others were concerned with listening attentively and trying to match the tones heard.
3. The next step was an attempt to teach the child to sing musical intervals. Animal ("cuckoo", "meeow", "hee-haw") and siren sounds were sung, and as the two-note patterns were learned at one pitch, they

were attempted in higher and lower pitch levels. The chime bars were again played by individuals while the group listened and echoed.

4. Additional intervals and short tunes were attempted using "lah" and "moo", then words. Short melodic patterns descending from the dominant and question/answer responses were also added to the singing repertoire. Individual work continued to attempt to extend the student's range, beyond A' if possible.

5. Two individual sessions were spent using a delayed feedback tape recorder. The child was instructed to listen, sing, and compare the sound heard through earphones with the note he or she was singing.

The development of concepts about pitch was not emphasized in the remedial program.

All students were pretested, and after eight weeks, were posttested with several tests of musical recognition (perception) and production (singing). Each recognition test was played on a piano and required a judgement from students as to whether the second of two notes or patterns was the same or different. The production tests attempted to measure students' ability to echo-sing:

- an individual pitch (single note production),
- a short phrase of 2 - 4 notes (interval production),
- a short tune (melody production),
- a rhythm pattern (rhythm production), and
- the child's choice of a song (free song).

A piano provided the stimulus pitches, and the production test was administered to students individually.

The analysis of variance with repeated measures used to analyze the data showed that the group that received remedial vocal training improved significantly more than the traditional group on two tests: single note production and interval production. There was no significant difference between the improvement of the two groups on the melody singing test or the free song test, nor on any of the recognition tests. This indicates that the remedial training was effective only on the tasks requiring simpler singing skills. The scores of the traditional and control groups were compared to evaluate the effects of extra instruction. Only the test of interval recognition showed a significantly greater change in scores for the traditional group, indicating that, for the most part, the influence of extra instruction was no greater than the influence of maturation in improving singing skills.

In this experiment, students were randomly selected and assigned to training, and the effects of maturation and extra instruction were controlled. The weaknesses were the low reliability of some of the tests used and the lack of consistency within the remedial teaching program. Additional factors which may explain the lack of overall improvement in singing and recognition skills are the lack of "normal" singers in the group training, and the lack of instruction in concepts about pitch.

Although each music education student taught remedial and traditional programs, differences in teaching style and lack of previous teaching experience would have caused random differences in the way both the remedial and the traditional programs would have been taught. The primary concern in this regard is whether all remedial groups would have received the same treatment, and it seems likely that ten different student teachers would not have delivered the same program. The use of observers would have clarified whether the remedial program was taught consistently by different teachers.

Test reliability is a very important aspect of experimental design. If the test does not accurately and consistently measure what it was designed to measure, the conclusions based on the test scores will be open to question. Roberts and Davies' 1975 recognition tests have low split-half reliabilities: single note recognition test r = .46, melody recognition test r = .21, interval recognition test r = .68. These values indicate that each test does not measure melodic perception in a consistent manner, and it would be unwise to make generalizations about pitch discrimination and tonal memory skills as measured by these tests. The singing tests have higher split-half reliabilities: single note production r = .88, interval production r = .91, melody production r = .76. The two tests that showed significant differences between the remedial group and the traditional group (single note production and interval production) had high enough reliabilities to consider the results dependable.

All group instruction in this study was conducted with students evaluated as "monotones" by their teachers. There were no normal singers to provide singing models for their classmates. Teachers in Gould's project noticed that when the poor singers began to improve their singing and recognize that they were beginning to sing in unison with the better singers in their class, their attitudes changed, they were very proud of their new skill, and they became eager to sing alone (Gould, 1968a, p. 26). The motivation to continue their accomplishments probably would not have occurred if the poor singers had not had the better singers to imitate. Although there is some controversy about the best type of vocal model for children (Goetze, 1990), the children in Grades 1 through 6 involved in Green's research (1987) were more accurate in pitch matching when they were responding to another child's singing, rather than to an adult female or male. In Roberts and Davies' study (1975, 1976), the poor pitch singers did not show a significant improvement on the tests of melody production and free song. These were more difficult tests than the single pitch test and interval test where improvements did occur. One reason for the lack of improvement may have been the lack of peer models. If normal singers were included in the training group with "monotone" singers, the poorer singers might improve more than if they were taught in a group with only inaccurate singers. The greater improvement should be reflected in significantly higher scores on more difficult singing tasks.

Although Roberts and Davies used Gould's project as a guide, the experimental program used in their study was a variation of Gould's experimental sequence. Gould stressed the importance of teaching both skills and concepts, but their remedial training program focused on vocal skill development and did not emphasize the development of concepts about pitch. This lack of training in pitch concepts may have contributed to the lack of improvement shown in pitch recognition. It also may have been a factor in the lack of improvement in melody-singing and the free song test. These more difficult singing skills probably require a better understanding of more complex pitch concepts than the simpler skills. Possibly the students were not able to improve in the more difficult singing skills because the required pitch concepts had not been introduced or reinforced. Roberts and Davies suggested that, because their program focused on voice production, one would expect recognition skills to improve less than singing skills. They reason that only the simpler singing tasks improved because those were the skills that were practiced.

Gould's program was used more recently as part of the instructional component in Kramer's research with Grade 3 and 4 students (1985). Kramer supplemented the Gould program with "music/imagery strategies" (p. 67) in a vocal setting, with the intent that these strategies would assist students in comprehending the singing process and help them to improve their singing accuracy. He stressed the importance of tonal images, and the kinesthetic sensations and accompanying kinesthetic images involved in accurate singing. Some of the music/imaging strategies employed in Kramer's experimental method were visual, physical and vocal "tracing" of a rainbow shape, and singing high and low progressions while showing the melodic shape with arm and hand motions. Similar techniques to kinesthetically and visually reinforce melody were suggested and used by several researchers previously mentioned in this review (Gould, 1968a; Apfelstadt, 1983; B. Jones, 1981; Richner, 1976). The experimental program was more effective than the activity-oriented curriculum used with the control group in improving singing accuracy, as measured by a shortened version of the Gould Speech and Song Response Test (Gould, 1968a). The difference between Kramer's experimental program and Gould's basic sequence was essentially one of emphasis. Kramer expanded on one component of Gould's program - the use of strategies designed to combine mental images of sound with images of kinesthetic (vocal) response - and spent a larger proportion of instructional time on these strategies than the teachers in Gould's experiment.

The results reported by Roberts and Davies (1975), Kramer (1985), and Gould (1968a), indicate that the basic approaches to singing improvement and the techniques suggested by Gould were effective, and were resilient to some degree of modification and adaptation.

In comparing the various teaching programs, it appears that more researchers had success with vocal training as a method of improving singing accuracy than melodic perception training alone. However, it appears that the most effective programs combined training in vocal control and melodic perception instruction in a sequenced program.

In many ways, Joyner's approach was quite similar to Wolner and Pyle's, since both programs included vocal training and pitch discrimination/pitch conceptual training. Although neither of these "experiments" had control groups, both appeared to be effective. Joyner began with pitch matching and proceeded to exercises to extend the students' comfortable range. Similarly, Wolner and Pyle's training sequence began with echo singing single pitches and then more difficult intervals, although they emphasized pitch discrimination training to a greater extent. The tonal, kinesthetic and visual imaging used by Kramer required students to create images of pitch changes and of their singing responses - essentially promoting a non-verbal understanding of pitch concepts concurrent with the development of vocal control. Gould also stressed the importance of training children to recognize higher/lower sounds (pitch discrimination) and develop concepts about pitch, in a vocal training program. Although students began with speech inflection exercises, the sequence continued with short restricted-range pattern echoes sung on a neutral syllable, gradually including longer melodic patterns with an increasingly wider range. Roberts and Davies' program was similar to Gould's, including a defined sequence of activities moving from speech to song. Forcucci's remedial program was also sequenced, beginning with pitch matching in a comfortable range, increasing the range gradually, and reinforcing high and low spatially on the (horizontal) keyboard. Richner's program was successful with Grade 5 students only. Possibly the reason for its lack of overall success was that only vocal exercises were taught; pitch concepts and a clear sequence were not included.

All of the effective programs included a sequence of vocal skills to be learned, beginning with speech inflection or pitches that could be produced easily, and proceeding gradually to extend the range of pitches sung and the length of melodic pattern (Gould, 1968a; Roberts and Davies, 1975; Kramer, 1985; Forcucci, 1975; Joyner, 1969; Wolner and Pyle, 1933; M. Jones, 1979).

All of these programs also included training to encourage perception of melodic intervals and to promote the understanding of concepts about high and low sounds and the similarity or difference of sounds and melodic patterns.

Programs that can be effectively taught in a group setting are the most practical and are, therefore, more likely to be attempted by music teachers. Gould's sequence was found to be effective with groups of children. Because Phase II and III had produced such positive results, the experimental sequence which had been developed at the end of the project might be effective in improving children's singing. Although the program was devised to help out-of-tune singers learn how to sing, it was observed that the singing of the other children in the class improved as well. Gould's sequence was designed for classroom use; it was based on the belief that all children could learn to sing in tune through group instruction if appropriate training was begun early enough.

Several components of Gould's and Roberts and Davies work are valuable. In order to further assess the effectiveness of the teaching techniques suggested, an experiment could be designed where the teaching program would be consistently taught, and reliable measuring instruments would be employed. Because the effects of extra small group practice were not found to significantly influence singing scores in Roberts and Davies' research, the teaching program could be taught in a regular class setting, if opportunities for individual singing during the group lesson were created. Developing pitch discrimination and learning concepts about pitch appear to be vital ingredients in the process of learning to sing; therefore, the teaching of pitch concepts should also be included. The final version of the Gould's experimental sequence had not been tested during his experiment; it was outlined in a manual called Finding and Learning to Use the Singing Voice (Gould, l968b). A modified version of this program was found to be successful with older (Grade

3 and 4) students by Kramer; however, the final sequence had not been tested with Grade 1 students prior to the present investigation. Grade 1 students appear to be ready for appropriate vocal instruction and melodic perception training. Petzold (1963) and Zwissler (1971) suggested that there is a need for these types of instruction to begin at the Grade 1 level. In addition to their developmental readiness, if Grade 1 students had the opportunity to develop vocal control and melodic perception skills in their class setting, the stigma of having to attend remedial classes at an older age would not be felt, and music learning would be a more positive experience.

CHAPTER III

METHOD

Research Design

A quasi-experimental design, the non-equivalent control group design (Campbell & Stanley, 1966), was chosen for this study. This design involves two groups, both of which are pretested and posttested, and it controls for the effects of maturation, history, testing and instrumentation (Cook & Campbell, 1979). Pre-existing classes of children were taught in this study because random selection of individual students was not feasible. Classes were randomly assigned to type of instruction, experimental or control. A control group was necessary in order that any changes in experimental group scores from pretest to posttest could not be attributed to the effects of maturation (changes in students' maturity), history (events occurring between pretest and posttest other than the instruction received), testing (the fact that students had been pretested), or instrumentation (any problems in the calibration of the tests or changes in the scorers). A pretest was given in order to determine the initial equivalence of groups. (See Figure 1 for the non-equivalent control group design.)

Subjects

Two schools in northwest Calgary, Alberta, in middle class suburban areas of the city, were selected for this study. Both schools, St. Bede School and St. Rita School, are in the Calgary Catholic School Board. Neither school had music specialists to teach Early Childhood Services (E. C. S.) or Grade 1 classes.

Figure 1 The Non-equivalent Control Group Design.

<u>Experimental Group</u>
Total n = 48 (n = 23, n = 25)

Selection:	Intact Grade 1 classes. Random assignment of class to type of instruction.
Pretest:	<u>Singing Achievement Measures</u>. <u>Primary Measures of Music Audiation</u>.
Instruction:	Music instruction with singing, movement, listening, and rhythm activities, including Gould's Speech to Song Sequence, taught by researcher.
Posttest:	<u>Singing Achievement Measures</u>. <u>Primary Measures of Music Audiation</u>.

<u>Control Group</u>
Total n = 52 (n = 27, n = 25)

Selection:	Intact Grade 1 classes. Random assignment of class to type of instruction.
Pretest:	<u>Singing Achievement Measures</u>. <u>Primary Measures of Music Audiation</u>.
Instruction:	Music instruction with singing, movement, listening, and rhythm activities, without Gould's Speech to Song Sequence, taught by researcher.
Posttest:	<u>Singing Achievement Measures</u>. <u>Primary Measures of Music Audiation</u>.

Four intact Grade 1 classes were used for the project, two in St. Bede School and two in St. Rita School. The school administrators attempted to balance the classes according to student ability and personality. In each school, one Grade 1 class was randomly assigned to experimental instruction, the other to control instruction, and classroom teachers were not informed as to the type of instruction received by their classes. The two classes that received experimental instruction were referred to as the Experimental Group, similarly the two control classes comprised the Control Group.

Because the dependent variables in the study required that subjects listen and respond to what they heard, any hearing problems that students had would hamper the validity of the pitch discrimination and singing test results. Hearing tests are given to students in the Calgary Catholic School Board only at the request of the teacher or the child's parents. No tests were requested for students involved in this study and the assumption was made that the hearing acuity of all participants was in the normal range.

A total of 105 students were enrolled in the four classes initially. The results of four students who moved out of the school district during the instructional period, and another student who was absent for more than half of the music lessons, could not be used. A total of 100 students participated in the study, 48 in the experimental group,and 52 in the control group. The characteristics of students in the sample were as follows:

	Number	Girls	Boys	Mean age (initial)	Age range
Experimental 1	23	8	15	6 yrs. 2 mos. (73.74 mos.)	67 - 87 mos
Experimental 2	25	8	17	6 yrs. 2 mos. (73.88 mos.)	67 - 87 mos.
Control 1	27	10	17	6 yrs. 3 mos. (75.30 mos.)	68 - 90 mos.
Control 2	25	6	19	6 yrs. 2 mos. (73.96 mos.)	67 - 88 mos.

There were considerably more boys in the study than girls, but because the balance was in favour of the boys in every class, the classes were considered comparable. The experimental group had 16 girls and 32 boys; the control group had 16 girls and 36 boys. The mean age of each class was near 6 years 2 months, and the age range of the entire sample was 5 years 7 months to 7 years 6 months.

The number of students whose Primary Measures of Music Audiation (Gordon, 1979a) test scores could be used in the data analysis was 89 because some students answered the test items according to a pattern, either alternating same/different responses or circling the same pair of faces for the entire test. Ten students indicated by pattern-marking that they did not understand the pretest directions, and one of these students also pattern-marked her posttest. Another control group student who had apparently understood the pretest, marked her posttest according to an alternating pattern and her results could not be used. The PMMA student sample was as follows:

	Number	Girls	Boys	Mean age	Age range
Experimental 1	21	7	14	6 yrs. 2 mos. (73.91 mos.)	67 - 87 mos
Experimental 2	23	8	15	6 yrs. 2 mos. (73.91 mos.)	67 - 87 mos.
Control 1	23	8	15	6 yrs. 4 mos. (75.91 mos.)	68 - 90 mos.
Control 2	22	6	16	6 yrs. 2 mos. (73.91 mos.)	67 - 88 mos.

Instruments

Two measuring instruments were required to test the hypotheses of this investigation. One measured singing achievement and the other measured melodic perception. In addition, a questionnaire was used to determine the child's home musical environment and previous music activities.

Singing Achievement Measures

The Singing Achievement Measures (SAM) was designed by the researcher to be used as a pretest and posttest of singing accuracy in this experiment. This test, which is administered to children individually, measures single pitch matching, melodic pattern matching, and the ability to sing a short song. Test items are on tape, students' responses were audio tape-recorded, and the testing procedure lasted about ten minutes for each child. (See Appendix B for the SAM.)

Primary Measures of Music Audiation

The "Tonal" part of the Primary Measures of Music Audiation (PMMA) by Edwin Gordon (1979a) was used as a pretest and posttest of melodic perception in this study. This standardized test is designed to be used with groups of Kindergarten to Grade 3 students. The "Tonal" section of PMMA contains 40 paired musical phrases on cassette tape. If the two musical phrases sound the same, the child draws a circle around the pair of faces which are the same on the answer sheet; if the two phrases sound different, the child draws a circle around the pair of faces that are different on the answer sheet. The test tape includes 12 minutes of listening time, and the test requires approximately 20 minutes of administration time. The split-half reliability for Gordon's Grade 1 standardization sample (which was representative of a culturally heterogeneous group) is reported in the test manual as $r = .89$, and the test-retest reliability for a two-week interval is reported as $r = .70$ for Grade 1 students.

Home Musical Environment Questionnaire

The Home Musical Environment Questionnaire (HMEQ) was used to respond to secondary hypotheses concerning the relationship of home musical environment to singing achievement and melodic perception. This self-reporting questionnaire for parents was designed by the researcher to determine the extent and quality of the child's previous and current experience with music outside the school and was completed by parents of all children involved in the project. (See Appendix C for HMEQ.)

Scoring Procedure. The scoring of the HMEQ consisted of converting the responses to numerical scores and totalling the scores in each category: Total, Siblings, Parents, Child. The "Total" score expressed the extent to which parents participated in music, the level of parental and family member involvement with the child in music, and the extent to which the child sang, listened, and played instruments alone and in organized activities. Siblings' participation in musical activities was in a separate category and was not included in the "Total" score because several participants did not have brothers and sisters. The "Parents" score measured the degree to which parents sang and helped the child to learn songs. The "Child" category was included to determine how supportive the child's environment was of the child's singing, and whether parents noticed their child singing at home. The "Parents" and "Child" categories consisted of responses that were included in "Total", but they were isolated in order to determine whether singing activities of parents with their child had a strong influence. The range of possible scores in each category was: Total 17 - 102; Siblings 4 - 24; Parents 2 - 12; Child 1 - 6.

Development of the Singing Achievement Measures

The tests used in previous related research were considered for use in this experiment. The "production" part of Roberts and Davies' test (1975), seemed from the descriptions given by the authors to be suitable for this experiment: it included 10 single pitch items, eight short melodic patterns, two longer patterns, and a song. Unfortunately, the test is not published in any articles, and attempts to obtain the dissertation containing it, or to obtain the test directly from the authors, proved fruitless. Gould's Speech and Song Response Test (1968a, p. B3 - B10) was considered to be too long and too difficult for Grade 1 students, and to include unnecessary items. To shorten the test by eliminating these items would mean that the test results could not be directly compared with Gould's results, and therefore, any advantage in using his test would be lost.

Characteristics of the Components of the SAM. The SAM is in three sections: Part I (single note echo singing), Part II (melodic echo singing), Part III (song singing). When designing the SAM, decisions were made regarding the following characteristics of the test items:

1. What proportion of the test should consist of single note, melodic pattern, and song items?
2. What singing range should be covered?
3. What features should the melodic patterns have?
   - what length should the patterns be - how many notes/beats?
   - in what tonality should the patterns be?
   - should the patterns be sung to neutral syllables or words?
4. What features should the song have?
   - should the same song be sung for pretest and posttest?
5. What type of vocal model should be used?

To assist the decision-making about the singing test, the literature was reviewed for relevant information pertaining to test design. The above questions will be addressed in the following paragraphs.

Researchers have employed various methods to measure singing accuracy. As evidence to their belief that the ability to echo sing a single pitch is important as the basis for replication of a phrase or a melody, several researchers measured single pitch matching (Jersild & Bienstock, 1934; Patrick, 1978; Welch, 1984). However, some music educators thought that a child singing one note out of context has little relationship to the child's ability to sing a melody (Boardman, 1964; Jordan-DeCarbo, 1982). Singing a melody is the desired outcome. Possibly the child is less motivated to sing one pitch accurately than to sing a song he or she likes. Hearing the note within a phrase or melody gives the note some musical meaning in relation to the other notes in the phrase. Joyner (1969) tested the ability of 11-year-olds to sing the entire British National Anthem, and Smith (1963) measured preschoolers' ability to sing lengthy phrases from known songs. In the SAM, single pitch matching, phrase singing, and song singing were all included so that different singing skills could be compared if desired. Because pattern singing is a more musical skill than single pitch singing, yet does not require previous knowledge of the melody the way that song singing does, the emphasis of the SAM is on echoing melodic patterns.

There is a marked lack of unanimity in the literature about the range of notes that children can sing, and this may be due to the use of different assessment methods. It could also be due to the wide individual variation in vocal range observed by Buckton (1977), and more recently by Flowers and Dunne-Sousa (1990). A range of approximately A to G' was observed by Hattwick (1933), Young (1971) and Buckton (1977). Other researchers observed higher or wider ranges. Greenberg (1979) suggested that most young children can sing comfortably in the D' to G' range, and that by age five, this range can expand to about C' to C". He recommended that parents try to sing in a light, high voice for their children to imitate. Jersild and Bienstock (1934) found that 50% of five-year-olds could sing A to D", and 50% of six-year-olds could sing A to G". Gordon (1971) suggested that songs be pitched in the D' to A' range.

For this study, it was assumed that the range of notes sung most easily by five- to seven-year-olds was D' to A'. Most of the melodic patterns in the Singing Achievement Measures were placed within this range so that children would not be frustrated in their attempts to echo sing the patterns. Lower notes were excluded because one goal of instruction was to extend the singing voice beyond the speaking range. If points were awarded for singing in the speaking range, singing achievement would not be reflected in the scores. Three patterns containing a few notes higher than A' (B' to D") were included in order to reflect the abilities of those children who could sing higher. In the single pitch matching part of the SAM, notes lower than D' were included in order to determine whether a weak singer a) could not match pitches at all (suggesting a perception problem), or b) could match only pitches lower than D'.

When considering the optimum phrase length for children's echo singing, Shuter-Dyson's discussion about perception (1968) was considered. She indicated that "it is not the number of notes that is important in musical perception, but the complexity of the ways in which they are classified and analyzed" (p. 201). It seemed that a sense of musical phrase was a better criterion for determining pattern length than the number of notes contained. To establish a sense of metric balance, and make it easier for children to perceive the patterns, all patterns had the same number of beats. All 10 patterns in SAM are four-beat phrases.

Jarjisian's 1981 results are pertinent to the choice of pattern tonality. She found that Grade 1 students who received instruction in both pentatonic and diatonic patterns were significantly better at rote singing than students who received instruction in only pentatonic or diatonic patterns. In the SAM, the range and difficulty level of the patterns was deemed more important than the tonality of the patterns. Pentatonic patterns and major and minor diatonic patterns were included in the SAM because these are the tonalities found in children's songs. A variety of what were expected to be somewhat familiar patterns was chosen.

Boardman (1964) and Apfelstadt (1983) suggested that the use of words with musical patterns on a vocal test should make the task more accessible to the young child, and Boardman's singing test contains musical phrases with words. Goetze (1985) found that Kindergarten and Grade 1 students sang more accurately with the syllable "loo" rather than text in individual and group situations. However, when Smale (1987) replicated Goetze' study with four- and five-year-olds, she found no significant difference between their ability to sing patterns to "loo" or to text. Sims, Moore and Kuhn (1982) also found no significant difference between the use of melismatic or syllable patterns in the number of pitches sung correctly by five- and six-year-olds. The literature is not conclusive about whether a neutral syllable or words have more effect on a child's ability to sing a pattern accurately. All single pitches and patterns in the Singing Achievement Measures were sung and echoed with "loo".

A requirement for the pretest song was that it should be familiar to Grade 1 students. The Farmer in the Dell was chosen, and in the pilot test this song was also used as the posttest song. Reports by Shuter-Dyson (1968, p. 205) and Roberts and Davies (1975) concur with Gould's hindsight (1968a), that when the same songs were used on pretest and posttest, "bad habits recurred on these songs which were not evident in songs learned after progress with singing skills had occurred" (Gould, 1968a, p. 16). In the main experiment, the pretest song was The Farmer in the Dell and the posttest song was Here we go Looby Loo. Both songs consisted of four phrases, had simple words, were in the same range and same meter (6/8). The posttest song was sung frequently during both the experimental and control instructional components and not reinforced through vocal exercises. Both songs were in the key of F major, and the range of each was C' to D".

Some researchers have found that when young children attempt to echo sing, they respond most accurately to a female vocal model (Sims, Moore, & Kuhn, 1982; Small & McCachern, 1983; Smith, 1963). More recently, Green (1987) reported that there were more correct responses to a child model than the female or male vocal model. The singing that children attempted to match on the SAM was that of a soprano with perfect pitch, a light tone quality, and a very slight vibrato typical of the adult trained voice.

Choice of Specific Melodic Material. The melodic patterns for the SAM were determined in consultation with Lois Choksy, based on her judgement of the frequency with which these patterns occur in English-language children's traditional songs (L. Choksy, personal communication, July 23, 1987). Two additional major mode patterns were added to the six suggested by Choksy. Although the minor tonality is less common in traditional children's songs, two minor patterns were retained in the final version of the test, since the minor mode is common in classical music of Western culture.

Initial Choice of Pattern Order. On most tests, the test items are placed in the order of increasing difficulty. The difficulty level of the test items on the SAM was based on the difficulty of the chosen tessitura, the pattern difficulty, whether the pattern contained leaps or changes in melodic direction, and the last note of the previous pattern. In some cases, the chosen tessitura of a test item was changed to make the item easier or more difficult. The patterns were placed in the order of their anticipated level of difficulty.

The Rating Scale. For Part I (single pitch matching), one point was scored for each pitch sung correctly. A short slide into the correct pitch scored one point. Each pitch was presented twice and each attempt was scored, so that the maximum score possible was 20 points.

In Part II (pattern matching), the rating scale for the students' echoed responses to patterns considered pitch level, interval direction, and interval size. "Interval direction" and "interval size" were used to describe components of melodic contour. The maintenance of pitch level referred to whether or not the notes in the student's response stayed in the same range of pitches as the original, not necessarily whether the response contained the same absolute pitches. The rating scale placed a greater value on the maintenance of pitch level over the maintenance of pattern contour. The purpose was to evaluate to what extent students could sing accurately in the desired pitch range (D' to A'). Scorers first evaluated whether the response maintained the pitch level of the original, then they decided to what extent interval size and interval direction were accurate. Each pattern was evaluated according to the following scale:

The PITCH LEVEL is maintained and -

5 points: - all pitches are correct.

4 points: - 1 or 2 small inaccuracies in interval size, although interval direction is correct.

3 points: - several small or 1 or 2 large inaccuracies in interval size, interval direction is the same.

2 points: - several small or 1 or 2 large inaccuracies in interval size, interval direction is not entirely the same as the original.

The PITCH LEVEL is not maintained and -

3 points: - interval size and interval direction are the same as original, i.e. the pattern is correctly transposed.

2 points: - 1 or 2 small inaccuracies in interval size, interval direction is the same.

1 point: - great inaccuracies in interval size, interval direction is not entirely the same as the original.

0 points: - response is not in a singing voice.

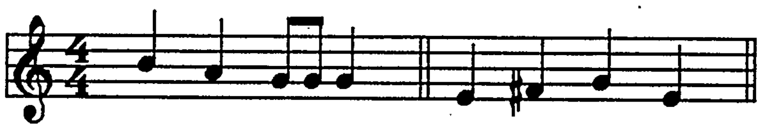
If the child did not sing the entire pattern, the pattern was rated as sung, and one point was deducted. If a repeated note was omitted, that is, if the rhythm was changed, a point was not subtracted. Each of the 10 patterns was presented twice, and sung twice by the students. The maximum possible score in Part II for 20 correct patterns was 100.

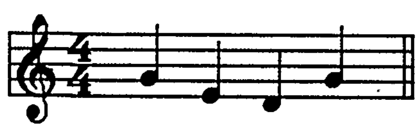
In scoring the song (Part III), scorers were instructed to listen for pitches only, not words. Each phrase in the song was rated according to the five-point scale outlined above. No points were awarded for beginning on the suggested starting pitch; pitch level was determined by the child's pitch level in the first phrase. The highest possible score in Part III was 20 points.

Creation of the Singing Test Tape. For the initial practice in administering the test, the investigator's own soprano singing voice was used, and was recorded using a Sony F-V30T microphone and Technics M229X tape deck with a Maxell CrO2 tape. The final version of the SAM employed the soprano voice described previously, using the same equipment. The tempo chosen for singing the patterns was not fast, so that the children had time to hear pitches, and an equal time was allotted between items for the child's response.

Preliminary Testing Procedures. When the first version of the SAM was completed, the investigator administered it to young relatives, and then to children at a daycare in northeast Calgary. This was to try different administration procedures, and also to determine the difficulty and appropriateness of the test items. The patterns were scored according to the initial rating scale.

Changes to the Test after Preliminary Administration. The items were ranked by difficulty according to the number of daycare children who scored 5 or 4 points on the item. This ranking was compared with the ranking by item difficulty found by Sinor (1984, p. 124) in her research, so that the final test order from easiest to most difficult could be established. It is interesting to note that the two easiest patterns for Sinor's five-year-olds were very similar to the patterns ranking first and third by the 5- to 8-year-olds at the daycare:



However, the pattern ranking second at the daycare:  was very similar except for its rhythm to the pattern that was ranked last (i.e. 46th) by Sinor:  The difference in pattern rhythm or children's age may have been important, since the range of these two patterns was similar: D' to G' at the daycare, and E' to A' in Sinor's project. Based on the daycare results and Sinor's results, additional adjustments were made to finalize the SAM: changes in item tessitura and item order, and one item substitution. (See Appendix D for preliminary versions of SAM and comparison of daycare and Sinor pattern rankings.)

The Testing Procedure. The administration of the singing test took place in a quiet room in each school: the reading specialist's room, an office, or a large storage room. The researcher walked to and from the testing room with each child, and administered the singing test to all children individually. In order to help students feel comfortable about the researcher and the testing procedure, conversation was made with each child on the way to the testing room, asking him or her whether songs were ever sung at school or at home, commenting on artwork displayed in the school hallway, asking their favourite colour, etc. In the testing room, the child was shown the equipment, and allowed to try the microphone. Then the student was told, "You will hear a woman singing on the tape, I would like you to try to sing back or echo the same thing that you hear. Try to make your voice sound just like the voice on the tape. There are no words, only 'loo'! Just listen first, and then sing." Before Part II, students were told that they would hear not just one sound, but little songs without words, and they were asked to copy the singing as before. In Part III, the researcher said "Let's sing <u>The Farmer in the Dell</u>" (or <u>Here we Go Looby Loo</u>), and sang the opening pitches of the song (C' F') with either the words "The farmer ---" or "Ready sing ---" and encouraged the child to begin. The researcher nodded or smiled, giving non-verbal approval to the child after every response. At the end of the test, each child was allowed to choose a reinforcement sticker to wear on his clothing or hand. The <u>SAM</u> master tape was played on a Sony CFS-3000 cassette tape player, while the Technics tape deck and Sony microphone previously described recorded the <u>SAM</u> tape and the child's singing. "High position" CrO2 cassette tapes were used. The child's assigned numerical subject code was spoken into the microphone by the investigator before the testing began.

<u>Pilot testing of the SAM</u>. The <u>SAM</u> was administered to Grade 1 students who were not involved in the main experiment, and one week later it was re-administered. Before the individual testing began, there was a ten-minute orientation session involving group singing and a brief demonstration of the microphone. Three scorers evaluated the singing tapes, and the resulting

scores were correlated to determine the degree of consistency between the scorers (inter-rater reliability). The scores from the first administration of the test were also correlated with the scores from the second administration of the test to determine the test-retest reliability of the Singing Achievement Measures.

Procedure

Early in September 1987, two training sessions were held with the scorers to discuss the rating scale and to practice scoring some sample tapes of children's singing.

On September 16, 1987, the Primary Measures of Music Audiation was pilot tested with Grade 1 students at a school in a suburban area of northwest Calgary. The school was in the Calgary Catholic School Board, and was similar to the schools where the main experiment was conducted. The purpose of this pilot testing was to familiarize the investigator with the test administration procedure. Some preliminary procedures were used to establish that the children understood the concept that two things can look the same or different, and that two tunes can sound the same or different, and to make sure that the students understood how to mark what they heard on the test paper.

On September 17 and 18, 1987, the Singing Achievement Measures was individually administered to 21 Grade 1 students at the same school where the PMMA was pilot tested. On September 24 and 25, 1987, one week after the first administration, the SAM was re-administered to the same Grade 1 students. The test was administered and scored as described previously in the Instruments section. No music instruction, other than that normally provided in the school, had occurred between the first and second administration of the SAM. The purpose of the pilot testing was to measure the degree of consistency between the three scorers (the inter-rater reliability), and to determine whether the test-retest reliability of the SAM was sufficiently high to warrant its use in the experiment.

Early in September 1987, parents of students involved in the experiment were sent letters requesting permission for their child to participate in the study, and were asked to complete the Home Musical Environment Questionnaire. (See Appendix E for Letters.of Permission)

All students were randomly assigned a numerical subject code which identified them on the tape for their SAM test.

After the reliability of the SAM had been tested, the orientation sessions at St. Bede School and St. Rita School began. On September 29, 1987, the investigator met with each of the four classes for a thirty-minute session comprised mostly of singing activities, to establish rapport with the students and to prepare them for the testing. Echo singing with "loo" was introduced, and the pretest song (The Farmer in the Dell) was sung by the group with the promise that the game would be played in a future music class. The microphone and taping equipment were demonstrated, and the children were reminded that when their turn came, they should try to echo or "copycat" the singing on the tape.

The Primary Measures of Music Audiation group pretest was administered to the four classes on the mornings of September 30 and October 1, 1987, and the individual pretesting of the Singing Achievement Measures began on October 2 and continued to October 9, 1987. The individual pretesting schedule was such that students in the same class came for the test at different times of day.

Beginning the week of October 13, 1987, music instruction began. All four classes received 30-minute music lessons three times a week (Tuesday, Wednesday, Thursday) for eight weeks. All classes had an equal number of morning and afternoon sessions, and every lesson was audio tape-recorded. The Schedule of Instruction is indicated in Figure 2.

Figure 2. Schedule of Instruction

Classes	Tuesday	Wednesday	Thursday
Weeks 1-4			
9:00-9:30	E1	C1	E1
9:30-10:00	C1	E1	C1
1:00-1:30	E2	C2	E2
1:30-2:00	C2	E2	C2
Weeks 5-8			
9:00-9:30	C2	E2	C2
9:30-10:00	E2	C2	E2
1:00-1:30	C1	E1	C1
1:30-2:00	E1	C1	E1

The Instructional Component

The experimental and control classes were taught by the researcher following a Grade 1 music curriculum that included singing songs, rhythmic and movement activities, and listening. The main difference in instruction was the inclusion of vocal training according to Gould's Experimental Sequence for 15 minutes of every experimental lesson, and the singing of additional songs reinforcing concepts in the curriculum for 15 minutes of every control group lesson. During the last three instructional weeks, in response to the control group's need to be challenged, the control lessons contained more rhythm work, including figuring out rhythm patterns from songs.

The 15-minute experimental component of each experimental group lesson was based entirely on the teaching sequence described by A. Oren Gould in Finding and Learning to Use the Singing Voice (1968b). This sequenced program began with exercises in speech inflection designed to develop an awareness in each child that he could control his speaking voice, and to establish a common vocabulary for high and low sounds. The inflection skills were transferred to the singing voice in a comfortable range using arm and hand motions to show high and low sounds. Short echo singing exercises sung to neutral syllables and animal and environmental imitations were used to expand the voice into a singing range. Increasingly longer phrases were sung to "loo", while continuing to show melodic contour with arms. As individual students demonstrated increased singing accuracy, they attempted longer phrases with words. In every experimental lesson, several children had opportunities to sing alone. Each child received individual teaching in the skills of one stage only when he or she could perform the skills learned at the previous stage. In order to determine at which stage individual students were working and to plan appropriate exercises for the next lesson, the researcher listened to individual singing on lesson tapes after the lesson and made notes in students' records of their singing progress. (See Appendix A for Gould's Speech to Song Sequence.)

In every experimental and control lesson, an attempt was made to include an action song, a traditional singing game in circle, line, or partner formation (Choksy & Brummit, 1987), and a listening song (ballad or lullaby) sung by the researcher. Both groups reinforced basic beat and rhythm concepts through clapping, tapping, marching, etc., and sang many of the same songs, but in the control lessons, rhythm concepts and skills were reviewed to a greater extent. (See Appendix E for Sample Lesson Plans.)

Instructional procedures were verified four times during the eight-week instructional component. Two graduate students in music education and two senior undergraduate students in music education each observed an experimental lesson and a control lesson and completed a form to record what they observed. The time spent in experimental vocal activities during the four observed experimental lessons was 13, 14, 14, and 17 minutes. None of these activities was observed in the control classes, and the remainder of each lesson was reported to have been taught the same way to both the experimental and the control groups. (See Appendix G for Observation Form.)

From December 4 to December 11, 1987, all students were individually posttested with the SAM by the investigator. The PMMA was administered to the four classes on the mornings of December 14 and 15, 1987.

The Home Musical Environment Questionnaire was evaluated and home musical environment scores were recorded for each student. Copies of all the pretest and posttest SAM tapes were made and given to the scorers, who did not know whether the tapes contained experimental group or control group, or pretest or posttest data.

Analysis of Data

Pilot test Singing Achievement Measures tapes were scored by the researcher and two other independent scorers. All scorers are musicians with graduate Kodály diplomas and experience with young children's singing. Inter-rater reliability of the SAM was determined, using the intraclass correlation coefficient procedure, statistic 10, in the Statistical Package for the Social Sciences x (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1983). The inter-rater reliability indicates the degree of consistency among scorers. The test-retest reliability of the SAM was calculated with the Pearson correlation coefficient, Pearson Corr in SPSSx. The test-retest correlation coefficient indicates the extent to which the test measures the same skills before and after the time interval.

The three scorers who rated children's singing in the pilot test, also evaluated the pretest and posttest Singing Achievement Measures tapes following the same rating scale. Each child's performance score was the mean of the raw scores assigned by the three scorers.

The significance level for all tests was set at $p = .05$. Pretest SAM performance scores from the two Grade 1 classes at School 1 were compared with the two Grade 1 classes at School 2 using a t-test to determine whether schools were equivalent on singing ability and melodic perception before instruction. The scores of the two experimental classes were combined to form the experimental group, and the scores of the two control classes were combined to form the control group for further data analysis. The pretest PMMA scores from the four classes were compared and combined in a similar manner.

To determine whether the experimental group and the control group could be considered equivalent in singing achievement before instruction, pretest SAM performance scores between the two groups (experimental and control) were tested for initial equivalence using a two-way analysis of variance. Factors were type of instruction and gender. Pretest PMMA scores of both groups were also compared using a two-way analysis of variance (type of instruction, gender). The MANOVA program in SPSSx was used for all analyses of variance and analyses of covariance in this study.

To determine the significance of any difference between the experimental group and the control group in singing accuracy after instruction, the posttest SAM performance scores of each group were compared using a two-way analysis of covariance with the pretest as the covariate. The second factor in the ANCOVA (gender) was included in order to determine whether boys or girls have significantly different singing accuracy (main effect for gender), or respond differently to treatment (instruction by gender interaction). This analysis was conducted separately for SAM Part I (single pitch matching), Part II (melodic pattern singing), and Part III (song).

To determine the significance of any difference between the experimental group and the control group in melodic perception after instruction, the posttest PMMA scores of each group were compared using a two-way analysis of covariance (instruction, gender) with the pretest as the covariate.

To compare the improvement from pretest to posttest in singing accuracy as measured by the SAM between the experimental group and the control group, an analysis of covariance with repeated measures and age as covariate was conducted on the pre- and posttest scores of both groups. This analysis was conducted three times, on Part I, Part II and Part III of the SAM. The effects of age were covaried out because of the evidence found by previous researchers that singing skills improve with maturation.

To compare the improvement from pretest to posttest in melodic perception as measured by the PMMA between the experimental group and the control group, an analysis of covariance with repeated measures and age as covariate was conducted on the pre- and posttest scores of both groups.

The Home Musical Environment Questionnaires were evaluated, and all students were assigned home musical environment scores in four categories: Total, Siblings, Parents, and Child. High, medium, and low rankings in home musical environment were established, the upper and lower quartiles being designated as high and low home musical environment respectively. T-tests were used to compare the singing and melodic perception skills (SAM and PMMA scores) of students from very good home musical environments (high HMEQ "Total" scores) with students from poor home musical environments (low HMEQ "Total" scores).

Results of all tests were correlated with all other tests using the Pearson Corr program in SPSSx. The correlations of primary interest were those between the HMEQ and SAM (home musical environment and singing accuracy), between the HMEQ and PMMA (home musical environment and melodic perception), and between the SAM and PMMA (singing accuracy and melodic perception).

CHAPTER IV

RESULTS

Pilot Test Results

Before the main experiment was conducted, the Singing Achievement Measures was pilot tested, and its inter-rater reliability and test-retest reliability were calculated to determine whether the test could be used dependably.

Inter-rater Reliability

The coefficients representing the correlations between the three scorers' ratings of students' singing on the Singing Achievement Measures are presented in Table 1. The coefficients ranged from r = 0.825 to r = 0.991 for the first administration of the SAM. The coefficients for the second administration, one week later, ranged from r = 0.845 to r = 0.994. These correlations represented high agreement among the three scorers, and indicated that the results could be subjected to further data analysis. The correlations compared quite favourably with the inter-rater reliabilities found by Boardman (1964), Sinor (1984), and Jarjisian (1981).

Table 1

Inter-rater Reliabilities for *Singing Achievement Measures* in Pilot Test

First Administration of SAM		Second Administration of SAM	
Part I	r = 0.991	Part I	r = 0.994
Part II	r = 0.982	Part II	r = 0.982
Part III	r = 0.825	Part III	r = 0.845

Test-retest Reliability

The correlation between students' performance scores on the first and second administrations of the Singing Achievement Measures Part I in the pilot test, was $r = 0.888$. The correlation coefficient between the first and second administrations of the SAM Part II was $r = 0.942$, and the correlation coefficient for Part III was $r = 0.414$.

The test-retest reliability represented by these coefficients for SAM Part I and SAM Part II indicated that the tests measured single pitch echo singing and melodic pattern singing consistently enough to warrant their use in the study. The relatively low correlation between the first and second administration of Part III, the song singing test, suggested that the test did not reliably measure students' accuracy in singing a song. Although the SAM Part III was retained as a measure of song singing in this study, its low reliability necessitated that its results be interpreted cautiously. (This is dealt with further in the Discussion.)

Results of Preliminary Data Analysis

The significance level was set at $p = .05$ for all data analysis in the study.

The inter-rater reliabilities for the SAM in the main experiment were $r = 0.991$ for SAM Part I, $r = 0.984$ for SAM Part II, and $r = 0.936$ for SAM Part III. Performance scores were calculated for each student, based on the average of the three scorers' rating for each score. (See Appendix H for the SAM Scorers' Ratings and the Performance Scores for all students. The raw scores for the PMMA pretest and posttest are included in Appendix I.)

Initial Equivalence of Classes

To determine whether the data from the two experimental classes could be combined, and whether the data from the two control classes could be combined, t-tests were performed comparing the two classes at

School 1 (E1 and C1) with the two classes at School 2 (E2 and C2). The results of these t-tests were:

SAM Part I	$t(98) = 0.96$,	$p = .341$
SAM Part II	$t(98) = 1.74$,	$p = .085$
SAM Part III	$t(98) = 2.03$,	$p = .045$
PMMA	$t(88) = -0.22$,	$p = .824$

There was no significant difference between schools on the initial scores of the SAM Part I, SAM Part II, and the PMMA. Therefore, for the main data analysis of these three tests, the scores from the experimental classes in both schools could be combined, and the scores from the control classes in both schools could be combined. These combined groups were referred to as the Experimental Group and the Control Group for the remainder of the analysis.

There was a significant difference between schools on the initial scores of SAM Part III. Therefore, the classes' scores could not be combined and the main data analysis of this test was performed on the data of the four classes rather than the combined groups.

Initial Equivalence of Groups

Before instruction, the experimental group had slightly higher scores in pattern singing, and melodic perception than the control group. The control group had higher pretest scores in single pitch singing. Table 2 presents the pretest means for SAM Part I, SAM Part II, and PMMA. This data was compared to determine to what extent the groups were initially equivalent using a two-way (instruction, gender) analysis of variance for each test. (See Tables 3, 4, and 5 for the pretest ANOVA.) The main effects for method of instruction were:

SAM Part I	$F(1,96) = 0.043$,	$p = .835$
SAM Part II	$F(1,96) = 0.006$,	$p = .939$
PMMA	$F(1,85) = 1.874$,	$p = .175$

Table 2

Pretest Means and Standard Deviations for *Singing Achievement Measures* Part I, Part II, and *Primary Measures of Music Audiation*

Group	SAM Part I mean	SD	SAM Part II mean	SD	PMMA mean	SD
Experimental	10.83	6.47	56.01	18.89	31.84	3.80
Control	11.14	6.82	54.72	20.03	31.09	3.78

Table 3

Analysis of Variance for *Singing Achievement Measures* Part I Pretest

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	4223.55941	96	43.99541		
Constant	11274.85685	1	11274.85685	256.27348	0.0
Instruction	1.90822	1	1.90822	.04337	.835
Gender	109.58537	1	109.58537	2.49084	.118
Instr. by Gender	.80235	1	.80235	.01824	.893

Table 4

Analysis of Variance for *Singing Achievement Measures* Part II Pretest

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	36597.28464	96	381.22171		
Constant	275162.26804	1	275162.26804	721.79065	0.0
Instruction	2.21803	1	2.21803	.00582	.939
Gender	500.50419	1	500.50419	1.31290	.255
Instr. by Gender	119.04508	1	119.04508	.31227	.578

Table 5

Analysis of Variance for *Primary Measures of Music Audiation* Pretest

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	1222.87149	85	14.38672		
Constant	77194.24599	1	77194.24599	5365.65858	0.0
Instruction	26.96332	1	26.96332	1.87418	.175
Gender	.06782	1	.06782	.00471	.945
Instr. by Gender	28.62162	1	28.62162	1.98945	.162

There was no significant difference between the Experimental and Control groups on either the SAM Part I, the SAM Part II, or the PMMA. Based on the results of these tests, the two groups were considered equivalent on single note echo singing, melodic pattern singing, and melodic perception, before instruction took place.

The means and standard deviations of the pretest SAM Part III scores for each of the four classes are presented in Table 6. To determine whether there was a significant difference among the four classes on the SAM Part III, a two-way (instruction, gender) analysis of variance was performed on the pretest scores. (See Table 7 for ANOVA on SAM Part III pretest.) The results of this analysis showed no significant differences among the four classes: $F(3,92) = 0.389$, $p = .762$. However, there was a significant ($p = .033$) instruction by gender interaction, indicating that the same gender group did not score higher in each class. Because this interaction was present, the pretest scores were not used as a covariate on the posttest. Figure 3 illustrates the interaction between instruction and gender before instruction.

Figure 3 SAM Part III Pretest Mean Scores by Gender and Class

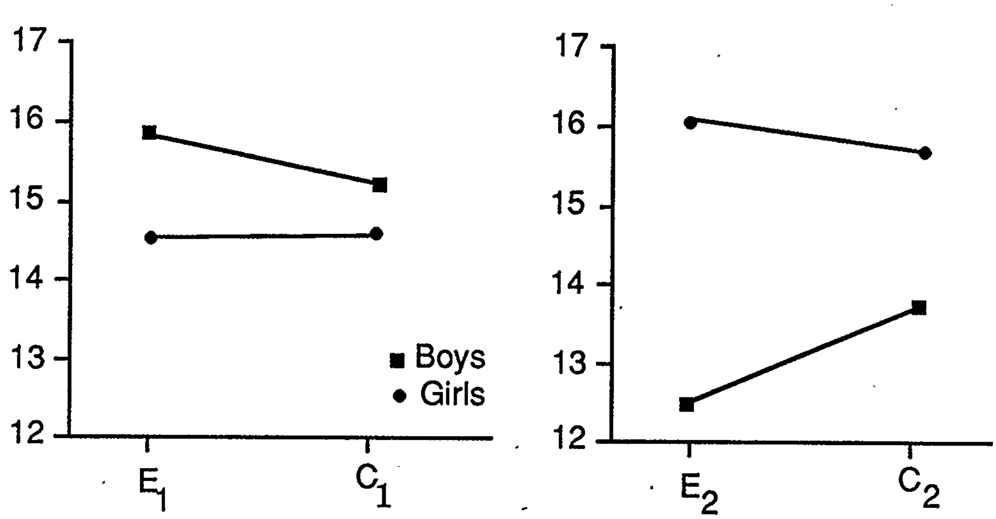


Table 6

Pretest Means and Standard Deviations for *Singing Achievement Measures* Part III

Group	mean	SD
Experimental 1	15.45	2.60
Control 1	15.04	2.32
Experimental 2	13.60	3.41
Control 2	14.27	4.17

Table 7

Analysis of Variance for *Singing Achievement Measures* Part III Pretest

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	881.62085	92	9.58284		
Constant	18628.30850	1	18628.30850	1943.92451	0.0
Instruction	11.16908	3	3.72303	.38851	.762
Gender	17.88523	1	17.88523	1.86638	.175
Instr. by Gender	87.35831	3	29.11944	3.03871	.033

Results for Main Hypotheses

Three separate two-way (instruction, gender) analyses of variance with pretest covariate were used to determine the significance of the difference between groups on SAM Part I, Part II, and PMMA data. The SAM Part III data was analyzed with a two-way (instruction, gender) analysis of variance (without pretest covariate).

The posttest means and standard deviations for the SAM Part I, Part II, and the PMMA are included in Table 8. The experimental group showed higher achievement in single pitch echo singing, melodic pattern singing and melodic perception than the control group, after instruction. The mean pretest and posttest scores for SAM Part I, SAM Part II, and PMMA are shown in Figure 4 Figure 5 and Figure 6, respectively.

The posttest means and standard deviations of the four classes on the SAM Part III are included in Table 9. The two experimental classes had higher posttest scores in song singing than the two control classes, after instruction.

The results related to the main hypotheses were as follows:

1. There is no significant difference in single pitch echo singing as measured by SAM Part I, between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

This hypothesis was rejected. $F(1,95) = 11.091$, $p = .001$

See Table 10 for ANCOVA for SAM Part I.

2. There is no significant difference in melodic pattern echo singing as measured by SAM Part II, between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

Table 8

Means and Standard Deviations for *Singing Achievement Measures* Part I, Part II, and *Primary Measures of Music Audiation*

Group	n	Pretest mean	Pretest SD	Posttest mean	Posttest SD
SAM Part I					
Experimental	48	10.83	6.47	14.88	5.54
Control	52	11.14	6.82	12.25	6.86
SAM Part II					
Experimental	48	56.01	18.89	70.12	17.29
Control	52	54.72	20.03	61.75	20.10
PMMA					
Experimental	44	31.84	3.80	35.25	3.14
Control	45	31.09	3.78	33.07	4.01

Table 9

Means and Standard Deviations for *Singing Achievement Measures* Part III

Group	n	Pretest mean	Pretest SD	Posttest mean	Posttest SD
Experimental 1	23	15.45	2.60	14.74	3.18
Experimental 2	25	13.60	3.41	13.88	3.55
Control 1	27	15.04	2.32	13.74	3.04
Control 2	25	14.27	4.17	12.52	3.71

Figure 4 SAM Part I Pretest and Posttest Mean Scores

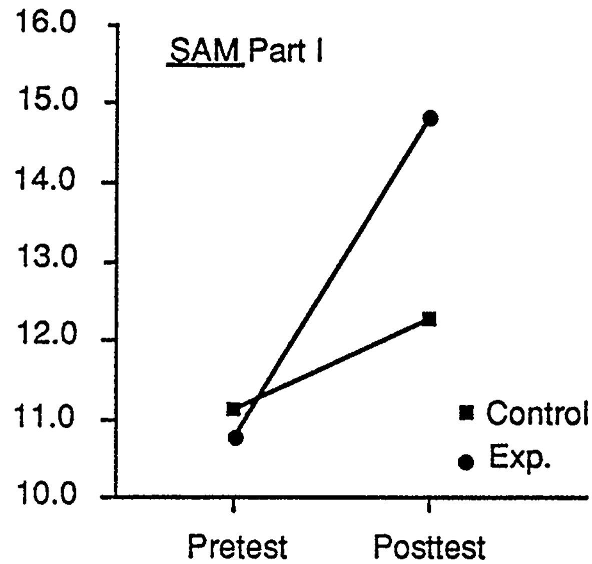


Figure 5 SAM Part II Pretest and Posttest Mean Scores

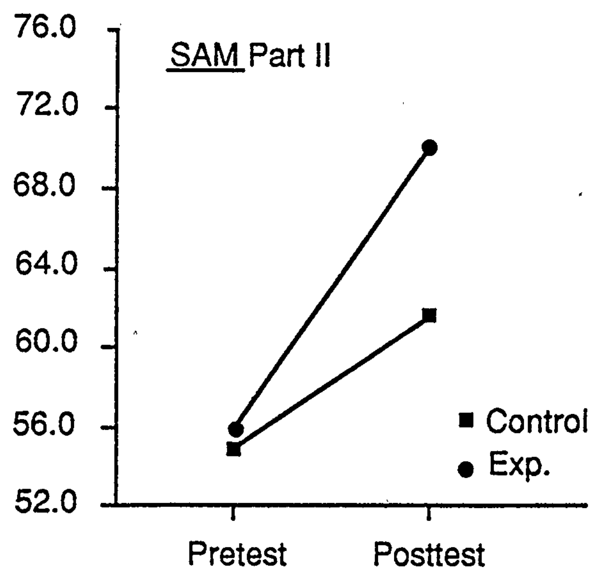


Figure 6 PMMA Pretest and Posttest Mean Scores

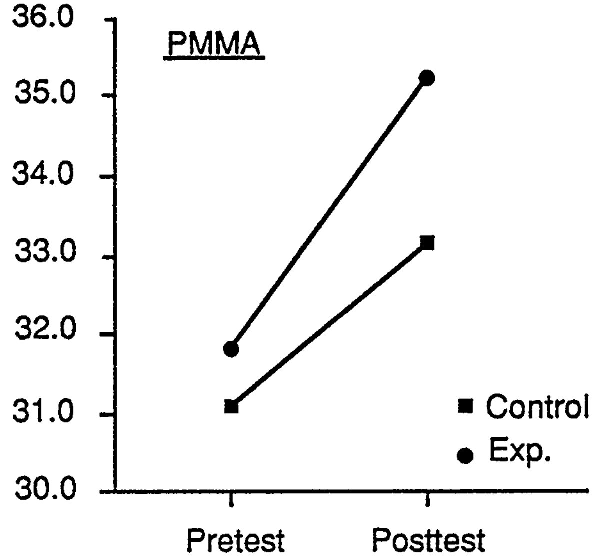


Table 10

Analysis of Covariance for *Singing Achievement Measures* Part I

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	1428.48619	95	15.03670		
Regression	2395.73296	1	2395.73296	159.32575	0.0
Constant	628.31971	1	628.31971	41.78575	0.0
Instruction	166.77492	1	166.77492	11.09119	.001
Gender	13.91079	1	13.91079	.92512	.339
Instr. by Gender	2.18594	1	2.18594	.14537	.704

This hypothesis was rejected. $F(1,95) = 15.772$, $p = .0001$

See Table 11 for ANCOVA for SAM Part II.

3. There is no significant difference in melodic accuracy when singing a song as measured by SAM Part III between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

This hypothesis was accepted. $F(3,92) = .731$, $p = .536$

See Table 12 for ANOVA for SAM Part III.

4. There is no significant difference in melodic perception as measured by the PMMA, between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

This hypothesis was rejected. $F(1,84) = 7.871$, $p = .006$

See Table 13 for ANCOVA for PMMA.

Alternative Data Analysis

The analysis of variance with repeated measures on one factor was used by Roberts and Davies (1975, p.232) in their study of singing improvement. In order that a comparison could be made between their results and the results of the present study, the repeated measures ANOVA was also used to analyze the present data. This analysis tests the significance of the difference between the improvement in each group from pretest to posttest. In addition, the effects of age differences on each dependent variable were covaried out. The mean scores for five-, six-, and seven-year-olds on the SAM Part I, Part II, Part III, and the PMMA are presented in Table 14. An examination of these means revealed that younger students tended to score lower than older students on all tests except the single note echo singing test (SAM Part I).

Table 11

Analysis of Covariance for *Singing Achievement Measures* Part II

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	7689.05236	95	80.93739		
Regression	26195.44057	1	26195.44057	323.65066	0.0
Constant	3843.74749	1	3843.74749	47.49038	0.0
Instruction	1276.54118	1	1276.54118	15.77196	.000
Gender	72.01074	1	72.01074	.88971	.348
Instr. by Gender	30.77131	1	30.77131	.38019	.539

Table 12

Analysis of Variance for *Singing Achievement Measures* Part III Posttest

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	1008.56965	92	10.96271		
Constant	16422.20880	1	16422.20880	1498.00582	0.0
Instruction	24.03223	3	8.01074	.73073	.536
Gender	7.55417	1	7.55417	.68908	.409
Instr. by Gender	81.86232	3	27.28744	2.48911	.065

Table 13

Analysis of Covariance for *Primary Measures of Music Audiation*

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	760.83772	84	9.05759		
Regression	346.13163	1	346.13163	38.21453	0.0
Constant	366.55700	1	366.55700	40.46959	0.0
Instruction	71.29072	1	71.29072	7.87082	.006
Gender	2.74457	1	2.74457	.30301	.583
Instr. by Gender	2.96730	1	2.96730	.32760	.569

Table 14

Five- Six- and Seven-year-olds' Mean Scores and Standard Deviations for *Singing Achievement Measures* Part I, Part II, Part III, and *Primary Measures of Music Audiation*

Age Group	n	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
SAM Part I					
Five-year-olds	34	9.18	6.81	11.38	6.62
Six-year-olds	58	12.12	6.34	14.68	5.87
Seven-year-olds	8	10.54	6.89	14.13	7.24
SAM Part II					
Five-year-olds	34	49.59	18.37	60.62	19.58
Six-year-olds	58	57.47	19.24	68.10	18.14
Seven-year-olds	8	64.38	20.66	70.71	22.60
SAM Part III					
Five-year-olds	34	13.90	2.81	12.95	2.97
Six-year-olds	58	14.83	3.47	14.00	3.60
Seven-year-olds	8	15.63	2.93	14.71	3.63
PMMA					
Five-year-olds	31	30.32	4.24	33.00	4.47
Six-year-olds	51	31.96	3.42	34.69	3.25
Seven-year-olds	7	32.86	3.49	35.29	2.80

Because the pretest and posttest SAM Part III were not the same, the repeated measures ANOVA was unsuitable; instead, a two-way (instruction, gender) analysis of covariance with age covariate was used.

The hypotheses tested, and the results of the two-factor (instruction, gender) ANCOVA with repeated measures on one factor and age covariate were:

5. There is no significant difference in improvement (pretest to posttest) in single pitch echo singing as measured by SAM Part I with age as a covariate, between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

This hypothesis was rejected. $F(1,96) = 10.009$, $p = .002$

See Table 15 for ANCOVA with repeated measures for SAM Part I.

6. There is no significant difference in improvement (pretest to posttest) in melodic pattern echo singing as measured by SAM Part II with age as a covariate, between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

This hypothesis was rejected. $F(1,96) = 14.140$, $p = .0001$

See Table 16 for ANCOVA with repeated measures for SAM Part II.

7. There is no significant difference after instruction in melodic accuracy when singing a song as measured by SAM Part III with age as a covariate, between the four Grade 1 classes.

This hypothesis was accepted. $F(3,91) = 1.926$, $p = .131$

See Table 17 for ANCOVA for SAM Part III posttest.

Table 15

Analysis of Covariance (Age Covariate) with Repeated Measures for *Singing Achievement Measures* Part I

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	842.92708	96	8.78049		
Time (pre - post)	237.73703	1	237.73703	27.07560	0.0
Instruction by Time	87.88797	1	87.88797	10.00946	.002
Gender by Time	20.23703	1	20.23703	2.30477	.132
Instr. by Gender by Time	1.44458	1	1.44458	.16452	.686

Table 16

Analysis of Covariance (Age Covariate) with Repeated Measures for *Singing Achievement Measures* Part II

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	4278.29854	96	44.56561		
Time (pre - post)	5018.83135	1	5018.83135	112.61669	0.0
Instruction by Time	630.14270	1	630.14270	14.13966	.000
Gender by Time	13.00118	1	13.00118	.29173	.590
Instr. by Gender by Time	26.18043	1	26.18043	.58746	.445

Table 17

Analysis of Covariance (Age Covariate)
for *Singing Achievement Measures* Part III Posttest

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	975.67361	91	10.72169		
Regression	32.89604	1	32.89604	3.06818	.083
Constant	13.17007	1	13.17007	1.22836	.271
Instruction	61.93556	3	20.64519	1.92555	.131
Gender	9.15196	1	9.15196	.85359	.358
Instr. by Gender	87.16256	3	29.05419	2.70985	.050

Table 18

Analysis of Covariance (Age Covariate) with Repeated Measures
for *Primary Measures of Music Audiation*

Source of variation	Sum of squares	*df*	Mean square	*F*	Signif. of *F*
Within cells	514.32492	85	6.05088		
Time (pre - post)	271.28903	1	271.28903	44.83463	0.0
Instruction by Time	18.64120	1	18.64120	3.08074	.083
Gender by Time	1.17788	1	1.17788	.19466	.660
Instr. by Gender by Time	.28957	1	.28957	.04786	.827

8. There is no significant difference in improvement in melodic perception as measured by the PMMA with age as a covariate, between Grade 1 students who received music instruction including Gould's Speech to Song Sequence and Grade 1 students who received music instruction without the activities in Gould's experimental sequence.

This hypothesis was accepted. $F(1,85) = 3.081$, $p = .083$

See Table 18 for ANCOVA with repeated measures for PMMA.

Results for Secondary Hypotheses

A secondary research question was whether there was a significant difference between boys and girls in singing accuracy or melodic perception, and whether boys or girls responded more favourably to the experimental instruction.

The pretest and posttest means and standard deviations for boys and girls separately are presented in Tables 19, 20, 21 and 22 for SAM Part I, SAM Part II, SAM Part III, and PMMA, respectively. For the entire sample, the girls' means were higher than the boys' means for all three parts of the SAM on both pretest and posttest. In the SAM Part III, the girls' mean was not higher in every class. The means of the PMMA indicated that for the entire sample boys had somewhat higher melodic perception scores than girls, on both pretest and posttest.

The results of all two-way (instruction, gender) analyses of variance indicated that there was no significant difference between boys' and girls' singing accuracy or melodic perception. There was a significant interaction between the method of instruction and the gender of students in two analyses: the SAM Part III pretest ANOVA, and the SAM Part III posttest ANCOVA with age covariate. Figures 3 and 7 illustrate the gender by instruction interaction.

Table 19

Boys' and Girls' Mean Scores and Standard Deviations for *Singing Achievement Measures* Part I

Gender	n	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
Girls	32	12.51	6.64	14.17	6.20
Boys	68	10.28	6.53	13.21	6.47
Experimental					
Girls	16	12.46	6.16	15.35	5.43
Boys	32	10.02	6.55	14.65	5.67
Control					
Girls	16	12.56	7.29	12.98	6.85
Boys	36	10.51	6.60	11.93	6.93

Table 20

Boys' and Girls' Mean Scores and Standard Deviations
for *Singing Achievement Measures* Part II

Gender	n	Pretest Mean	SD	Posttest Mean	SD
Girls	32	58.66	19.52	69.95	18.46
Boys	68	53.28	6.53	63.80	19.32
Experimental					
Girls	16	57.65	17.97	73.52	15.82
Boys	32	55.19	19.56	68.42	17.99
Control					
Girls	16	59.67	21.51	66.38	20.67
Boys	36	52.53	19.24	59.69	19.78

Table 21

Boys' and Girls' Mean Scores and Standard Deviations for *Singing Achievement Measures* Part III

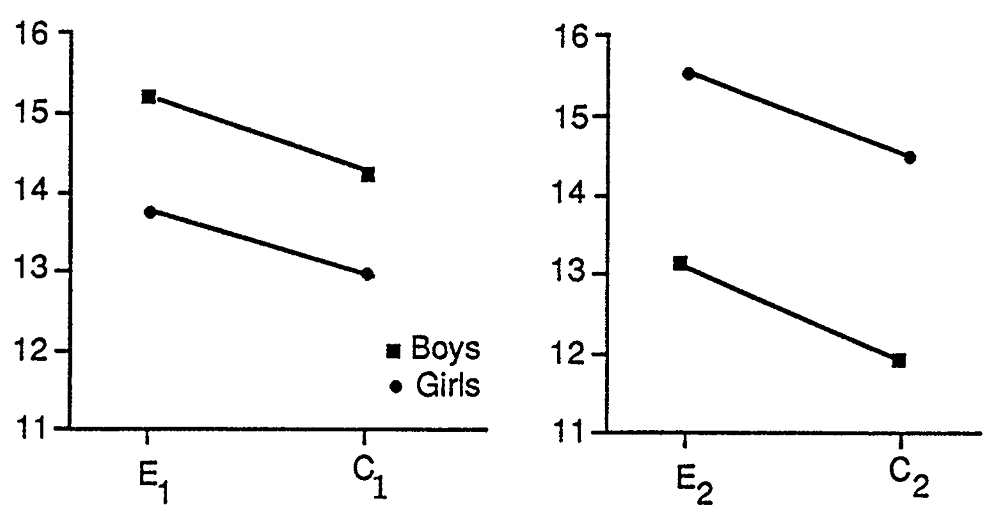
Gender	n	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
Girls	32	15.20	2.60	14.10	3.42
Boys	68	14.29	3.47	13.51	3.43
Experimental					
Girls E1	8	14.58	2.16	13.83	3.37
Boys E1	15	15.91	2.76	15.22	3.08
Girls E2	8	16.08	2.68	15.58	2.88
Boys E2	17	12.43	3.12	13.08	3.63
Control					
Girls C1	10	14.63	1.89	12.90	2.76
Boys C1	17	15.28	2.57	14.24	3.17
Girls C2	6	15.78	3.99	14.50	4.95
Boys C2	19	13.79	4.21	11.90	3.13

Table 22

Boys' and Girls' Mean Scores and Standard Deviations for *Primary Measures of Music Audiation*

Gender	n	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD
Girls	29	31.45	3.69	33.93	3.60
Boys	60	31.47	3.87	34.25	3.85
Experimental					
Girls	15	32.60	3.22	35.67	2.16
Boys	29	31.45	4.07	35.03	3.56
Control					
Girls	14	30.21	3.87	32.07	3.95
Boys	31	31.48	3.74	33.52	4.02

Figure 7 SAM Part III Posttest Mean Scores by Gender and Class



The results of the secondary hypotheses concerning gender were:

1. a) There is no significant difference between girls or boys in single pitch echo singing, as measured by the SAM Part I.

This hypothesis was accepted for both pretest and posttest.

Pretest: $F(1,96) = 2.491$, $p = .118$

Posttest: $F(1,95) = 0.925$, $p = .339$

1. b) There is no significant interaction between method of instruction and gender of subjects in single pitch echo singing, as measured by the SAM Part I.

This hypothesis was accepted for both pretest and posttest.

Pretest: $F(1,96) = 0.018$, $p = .893$

Posttest: $F(1,95) = 0.145$, $p = .704$

See Table 3 for ANOVA for SAM Part I pretest.

See Table 10 for ANCOVA for SAM Part I posttest.

2. a) There is no significant difference between girls or boys in melodic pattern echo singing, as measured by the SAM Part II.

This hypothesis was accepted for both pretest and posttest.

Pretest: $F(1,96) = 1.313$, $p = .255$

Posttest: $F(1,95) = 0.890$, $p = .348$

2. b) There is no significant interaction between method of instruction and gender of subjects in melodic pattern singing, as measured by the SAM Part II.

This hypothesis was accepted for both pretest and posttest.

Pretest: $F(1,96) = 0.312$, $p = .578$

Posttest: $F(1,95) = 0.380$, $p = .539$

See Table 4 for ANOVA for SAM Part II pretest.

See Table 11 for ANCOVA for SAM Part II posttest.

3. a) There is no significant difference between girls or boys in song singing accuracy, as measured by the SAM Part III.

This hypothesis was accepted for both pretest and posttest.

Pretest: $F(1,92) = 1.866$, $p = .175$

Posttest: $F(1,92) = 0.689$, $p = .409$

3. b) There is no significant interaction between method of instruction and gender of subjects in song singing accuracy, as measured by the SAM Part III.

This hypothesis was rejected on the pretest and accepted on the posttest.

Pretest: $F(3,92) = 3.039$, $p = .033$

Posttest: $F(3,92) = 2.489$, $p = .065$

See Table 7 for ANOVA for SAM Part III pretest.

See Table 12 for ANOVA for SAM Part III posttest.

Figure 3 shows the pretest interaction between instruction and gender and Figure 7 shows the posttest interaction (significant at $p = .10$ level) between instruction and gender.

4. a) There is no significant difference between girls or boys in melodic perception, as measured by the PMMA.

This hypothesis was accepted for both pretest and posttest.

Pretest: $F(1,85) = 0.005$, $p = .945$

Posttest: $F(1,84) = 0.303$, $p = .583$

4. b) There is no significant interaction between method of instruction and gender of subjects in melodic perception, as measured by the PMMA.

This hypothesis was accepted for both pretest and posttest.

Pretest: $F(1,85) = 1.989$, $p = .162$

Posttest: $F(1,84) = 0.328$, $p = .569$

See Table 5 for ANOVA for PMMA pretest.

See Table 13 for ANCOVA for PMMA posttest.

Additional Results

Having found that the experimental instruction was effective in improving singing accuracy, the question was considered whether all experimental group students benefitted equally or whether the improvement was concentrated in certain subgroups. To address this question, the gain or improvement in singing accuracy (as measured by the posttest-pretest difference on SAM Part II) of subgroups within the experimental group was compared with the gain in their corresponding control subgroup. The SAM Part II was chosen as the measure of singing accuracy with which to calculate gain scores, because it was the most reliable singing test used, and because it measured a more musical task than the single note singing test. Because raw gain scores are subject to regression, these results cannot be interpreted as conclusively as the ANCOVA analyses. All analyses using gain scores were conducted with the Statview program on an Apple MacIntosh computer.

Of particular interest was whether the gain in singing accuracy made by students who were weak singers before instruction took place, was different for students who received experimental or control instruction. The composite singing score which determined initial singing accuracy was the sum of the SAM Part I and SAM Part II pretest scores. Students whose composite singing scores were in the lowest 25% of the entire sample were considered to be weak singers. The mean gains on the SAM Part II made by weak singers in the experimental and control groups were compared with a t-test, which found there was a significant difference in the singing improvement made by weak singers who received the experimental instruction, and weak singers who did not receive the experimental instruction ($p = .01$). (See Table 23 for the results of this t-test.)

Table 23

T-test for Mean Gains on *Singing Achievement Measures Part II* between Experimental Group Poor Singers and Control Group Poor Singers

Dependent Variable	Group	n	Mean	SD	*t*	*df*	*p*
SAM Part II gains							
	Exp. poor singers	11	18.4	11.2	2.56	23	.01
	Cont. poor singers	14	7.3	10.5			

Table 24

T-test for Mean Gains on *Singing Achievement Measures Part II* between Experimental Students with Low Melodic Perception and Control Students with Low Melodic Perception

Dependent Variable	Group	n	Mean	SD	*t*	*df*	*p*
SAM Part II gains							
	Exp. low PMMA	7	12.0	8.1	0.57	19	.375
	Cont. low PMMA	14	9.7	8.9			

To determine whether students with low melodic perception who received the experimental instruction improved their singing accuracy more than students with low melodic perception in the control group, a t-test was conducted comparing the gain scores on the SAM Part II, of experimental students with low PMMA scores and control students with low PMMA scores. The low category of melodic perception was determined as the lower 25% of PMMA pretest scores. No significant difference was found at the p = .05 level, between experimental and control students in the low category of melodic perception (Table 24). Since the entire experimental group was found to have improved significantly more in singing accuracy than the entire control group on the repeated measures ANCOVA test, these results suggest that subgroups of students other than those with low melodic perception benefitted more from the experimental instruction.

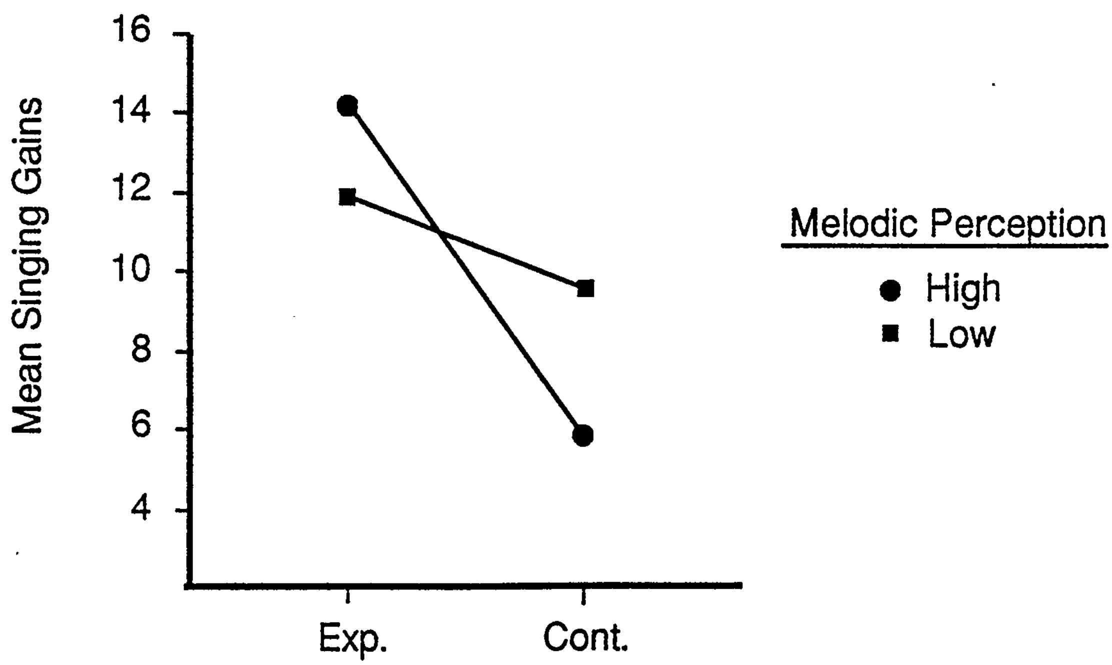
To determine whether students with high melodic perception who received the experimental instruction improved their singing accuracy more than students with high melodic perception in the control group, a t-test was conducted comparing the gain scores on the SAM Part II of experimental students with high PMMA scores and control students with high PMMA scores. The high category of melodic perception was determined as the upper 25% of PMMA pretest scores. The significant difference (p = .025) found in singing accuracy gain between experimental students with high melodic perception and control students with high melodic perception, indicated that the experimental instruction helped students with high melodic perception to improve their singing accuracy. (See Table 25 for t-test results.) As shown in Figure 8, the influence of high and low melodic perception on gains in singing accuracy was different for experimental and control instruction. With experimental instruction, the students with high melodic perception had greater

Table 25

T-test for Mean Gains on *Singing Achievement Measures Part II* between Experimental Students with High Melodic Perception and Control Students with High Melodic Perception

Dependent Variable	Group	n	Mean	SD	*t*	*df*	*p*
SAM Part II gains							
	Exp. high PMMA	11	14.1	9.2	2.17	19	.025
	Cont. high PMMA	10	5.9	8.1			

Figure 8 Mean Singing Gains on SAM Part II for High and Low Melodic Perception Students in Experimental and Control Groups



gains, but with control instruction, the low melodic perception students had greater singing gains.

Secondary Comparisons

All students' results on all tests were correlated with all other tests. (The correlation coefficients between all test results are included in Appendix J.) The comparisons of primary interest are: melodic perception and singing accuracy, home musical environment and singing accuracy, and home musical environment and melodic perception.

Melodic Perception and Singing Accuracy

To determine the degree of relationship between singing accuracy and melodic perception, the results of the single note echo singing, pattern singing, and song singing tests were correlated with the results of the melodic perception test. The coefficients representing the correlations between the results of Singing Achievement Measures Part I, II, and III, and the results of the Primary Measures of Music Audiation are presented in Table 26. The coefficients ranged from $r = 0.26$ to $r = 0.52$, indicating a moderately weak relationship between singing accuracy and melodic perception. The only coefficient higher than $r = 0.50$ represented the correlation between the posttest SAM Part II (pattern singing) and the PMMA posttest. In general, the highest correlations (comparatively) were between SAM Part II and PMMA on pre- and posttest in both measures.

To determine whether students with high or low melodic perception were significantly different in their singing accuracy, the pretest means on the SAM Part II were compared between the high and low melodic perception groups using a t-test. High and low melodic perception groups were the same as previously designated - the highest-scoring 25% and lowest-scoring 25% of students on the PMMA. The results showed that there was a significant

Table 26

Correlation Coefficients for *Primary Measures of Music Audiation* and *Singing Achievement Measures* Part I, Part II, Part III

	PMMA pretest (n)	PMMA posttest (n)
SAM Part I		
pretest	r = 0.45 (90)	r = 0.37 (98)
posttest	r = 0.44 (90)	r = 0.45 (98)
SAM Part II		
pretest	r = 0.48 (90)	r = 0.46 (98)
posttest	r = 0.46 (90)	r = 0.52 (98)
SAM Part III		
pretest	r = 0.26 (90)	r = 0.32 (98)
posttest	r = 0.36 (90)	r = 0.41 (98)

Table 27

T-test for *Singing Achievement Measures* Part II Pretest between Students with Low Melodic Perception and Students with High Melodic Perception

Dependent Variable	Group	n	Mean	SD	*t*	*df*	*p*
SAM Part II pretest							
	low PMMA	21	48.4	19.6	3.58	40	.0005
	high PMMA	21	69.0	17.7			

difference in pretest singing accuracy between the high melodic perception group and the low melodic perception group ($p = .0005$). (See Table 27 for the t-test results.)

Home Music Environment and Singing Accuracy

The mean score in home musical environment as determined by the Home Musical Environment Questionnaire (Total) for the entire sample was 54.93 and the standard deviation was 10.80. The possible range of scores in the Total category was 17 to 102, and the range of scores in the sample was 20 to 79.

The correlations for the entire sample between home musical environment in four categories and achievement in singing accuracy and melodic perception are presented in Table 28. In the Total and Siblings categories of home musical environment, the correlation coefficients ranged from $r = 0.06$ to $r = 0.21$ and indicated a low correlation between home musical environment and singing accuracy or melodic perception as measured by SAM Part I, II, III and PMMA.

To determine whether students from good home musical environments had achieved significantly higher skills in singing and melodic perception than students from poor home musical environments, t-tests were conducted comparing the SAM Part I, II, III and PMMA pretest means, between students in the high home musical environment group with students in the low home musical environment group. The highest-scoring 25% of students were designated as having high home musical environment, and the lowest-scoring 25% as having low home musical environment. The t-test results showed that students with high HMEQ scores and students with low HMEQ scores were significantly different in single note echo singing, melodic pattern singing, and

Table 28

Correlation Coefficients for *Home Musical Environment Questionnaire* and *Singing Achievement Measures* Part I, II, III, *Primary Measures of Music Audiation*

	HMEQ Total (n)	HMEQ Siblings (n)	HMEQ Parents (n)	HMEQ Child (n)
SAM Part I				
pretest	r = 0.16 (100)	r = 0.06 (91)	r = - 0.01 (100)	r = - 0.01 (98)
posttest	r = 0.17 (100)	r = 0.07 (91)	r = 0.02 (100)	r = 0.06 (98)
SAM Part II				
pretest	r = 0.17 (100)	r = 0.12 (91)	r = - 0.03 (100)	r = 0.06 (98)
posttest	r = 0.15 (100)	r = 0.15 (91)	r = - 0.08 (100)	r = 0.07 (98)
SAM Part III				
pretest	r = 0.18 (98)	r = 0.12 (100)	r = 0.01 (91)	r = 0.20 (100)
posttest	r = 0.21 (100)	r = 0.19 (91)	r = - 0.02 (100)	r = 0.19 (98)
PMMA				
pretest	r = 0.13 (90)	r = 0.18 (82)	r = 0.02 (90)	r = - 0.10 (88)
posttest	r = 0.14 (98)	r = 0.09 (89)	r = - 0.10 (98)	r = 0.06 (96)

song singing accuracy as measured by SAM Parts I, II and III. The difference between the melodic perception skills as measured by the PMMA of the two home musical environment groups was not found to be significant. (See Table 29 for the t-test results.)

Although home musical environment appeared to be a factor in singing accuracy as indicated by the t-test results, the low correlation coefficients suggested that the relationship was not direct, and that factors other than home musical environment had a stronger influence in the development of singing and melodic perception skills.

Table 29

T-tests for *Singing Achievement Measures* Part I, II, and III and *Primary Measures of Music Audiation* between High and Low Home Musical Environment Groups

Dependent Variable	Group	n	Mean	SD	Pooled Variance Estimate *t*	*df*	2 - tailed Probability
SAM Part I							
	low HMEQ	24	10.0	6.3	-2.08	44	.043
	high HMEQ	22	13.7	5.9			
SAM Part II							
	low HMEQ	24	50.6	18.7	-3.03	44	.004
	high HMEQ	22	66.6	17.1			
SAM Part III							
	low HMEQ	24	13.4	3.1	-2.68	44	.010
	high HMEQ	22	15.7	2.8			
PMMA							
	low HMEQ	21	30.9	3.6	-1.09	38	.281
	high HMEQ	19	32.2	4.0			

CHAPTER V

SUMMARY, DISCUSSION, AND RECOMMENDATIONS

Summary

This study was designed to assess whether the singing accuracy and melodic perception of Grade 1 students would improve significantly if they received music instruction containing a specific sequence of vocal skills and concepts. The design used compared the singing accuracy in three components (single note echo singing, melodic pattern singing, song singing) and the melodic perception of an experimental group and a control group after instruction.

The hypothesis was that the group receiving music instruction including the experimental sequence of vocal skills and concepts would achieve significantly better singing accuracy and melodic perception skills than the group receiving music instruction including the singing of songs without the experimental sequence. The null hypothesis was that there is no difference in the singing accuracy or melodic perception of Grade 1 students who receive music instruction that includes the Speech to Song Sequence of A. Oren Gould, and Grade 1 students who receive music instruction that does not contain the Gould sequence.

In each of two similar schools in Calgary, Alberta, two Grade 1 classes were selected and randomly assigned to either experimental or control instruction. For eight weeks, all four classes received 30-minute music lessons thrice-weekly, taught by the researcher. For 15 minutes of every lesson, all classes received instruction following the same Grade 1 music curriculum. For the other 15 minutes of every lesson, the experimental classes received sequenced instruction based on the gradual acquisition of vocal skills and melodic concepts, while the control classes received a traditional program, singing additional songs that reinforced concepts in the curriculum.

All students were pre- and posttested with the Singing Achievement Measures (SAM) and the Primary Measures of Music Audiation (PMMA). Parents of all students completed the Home Musical Environment Questionnaire, designed by the researcher to determine the extent of musical activities and influences in the child's home. The SAM was administered to students individually and was designed by the researcher to measure three components of singing accuracy - single note echo singing (Part I), melodic pattern echo singing (Part II), and song singing (Part III). Its test-retest reliability was determined through pilot-testing to be $r = 0.888$, $r = 0.942$, and $r = 0.414$ for the three parts.

A t-test determined that the two schools were initially equivalent in single note singing, melodic pattern singing, and melodic perception, and the data from the two experimental classes was combined and referred to as the experimental group for these three tests (SAM Part I, SAM Part II, PMMA). The control classes' data was similarly combined. Schools were not initially equivalent on song singing test scores, and the data from the four classes was analyzed separately for SAM Part III.

Previous research had indicated the possibility of differential skills for boys and girls, therefore gender of students was included as a factor in the data analysis. A two-way (instruction, gender) analysis of covariance with pretest covariate was used to determine the significance of the difference between group scores on SAM Part I, SAM Part II, and PMMA after instruction. A two-way (instruction, gender) analysis of variance was used to analyze the SAM Part III posttest data. The results indicated that the experimental instruction had a highly significant effect on the single pitch singing, pattern singing, and melodic perception of Grade 1 students. However, there was no significant difference in song singing as determined by the SAM Part III posttest scores, among the four classes.

An alternative data analysis was used to compare the improvement between control and experimental groups on SAM Part I, SAM Part II, and PMMA after instruction. Since the age range of students in the study spanned two years, from 5 years 7 months to 7 years 6 months, the effects of age on the singing accuracy and melodic perception scores were covaried out in this analysis. The results of the two-way (instruction, gender) analysis of covariance with repeated measures on one factor (instruction) with age as covariate showed that the experimental instruction had a highly significant effect on the improvement in single note singing and melodic pattern singing, but not on the improvement in melodic perception of Grade 1 students in the study. The results of the two-way (instruction, gender) analysis of covariance with age covariate which was used to analyze the SAM Part III posttest data, showed that there was no difference among the four classes in song singing accuracy after instruction.

Based on the Home Musical Environment Questionnaire data, students were assigned to high, medium, or low home musical environment categories. A t-test, which compared the SAM scores of students in the high and low home musical environment groups, indicated that there was a significant difference in singing accuracy between students from musical home environments and students having few musical influences in their home environment. T-tests comparing the PMMA scores of students with strong and weak home musical environmental influences showed that the difference between the high and low home musical environment groups in melodic perception was not significant.

Low correlations were found between home musical environment and singing accuracy, and between home musical environment and melodic perception, and moderate correlations were found between singing accuracy and melodic perception.

Discussion of Results

Main Hypotheses

Based on the results of this study, the null hypothesis was rejected and it was concluded that Grade 1 students who received music instruction including Gould's Speech to Song Sequence of vocal skills and concepts achieved more accurate single note echo singing, melodic pattern singing, and melodic perception than Grade 1 students who received music instruction including the singing of additional songs, without the Speech to Song Sequence. The difference between experimental group scores and control group scores was found to be highly significant on the posttests Singing Achievement Measures Part I and Part II, and Primary Measures of Music Audiation. In addition, a highly significant difference in improvement from pretest to posttest was found between group scores on the Singing Achievement Measures Part I and Part II.

Clearly, the experimental program was effective in improving singing accuracy in Grade 1 students.The components of this program that appeared to the researcher to be beneficial were:

1) the instruction in vocal skills concurrent with the teaching of concepts about pitch,
2) the sequenced aspect of the program, where new skills and concepts were developed from previously learned skills and concepts,
3) the individualized programming, where students received individual instruction within the group situation at their current achievement level or stage in the sequence, enabling them to develop confidence and a sense of

success about their skills,

4) the monitoring of each child's singing achievement by the teacher, and the careful planning of exercises appropriate for individual students,
5) the opportunity for the initial exploration of a variety of vocal pitches, followed by attempts at more precise imitation of narrower intervals,
6) the use of phrases sung to neutral syllables before the use of words, when students were echo singing individually,
7) the emphasis on careful listening by students when responding individually, and by the group when deciding whether two sung phrases were the same or different,
8) the opportunity for good singers to act as vocal models for their peers,
9) the combination of the sequenced program, and a program with broader goals that included activities such as traditional singing games, songs with actions, and listening songs.

In addition, the receptiveness of Grade 1 students to the program may have contributed to its success. Any initial inhibition some students may have had about singing individually was overcome quickly during the speech inflection exercises which were enjoyable and appeared to give the students a sense of success. Students were able to hear improvements in their classmates' singing and were very supportive of each other. This sense of classroom community was reinforced through the group singing games that were part of the "non-experimental" component of each lesson.

<u>SAM Part III Results</u>. Although the experimental instruction was effective in improving single note echo singing and melodic pattern echo singing, the difference between the experimental and control group scores in song singing as measured by the <u>Singing Achievement Measures</u> Part III, was found to be not significant.

The interpretation of the SAM Part III results is somewhat problematic due to the low test-retest reliability of this instrument (r = 0.414), as determined in the pilot test. The inter-rater reliability was found to be quite satisfactory on both the first and second administrations during pilot-testing (r = 0.825 and r = 0.845). This indicated that the scorers' use of the rating scale was consistent, and that the rating scale itself was reliable. However, the low correlation between the first administration and the second administration of the SAM Part III during the pilot-testing indicated that the test did not measure the same skills before and after the time interval, although the test song remained the same. An examination of the pilot test tapes revealed inconsistencies in the procedures used for tonally preparing the song in SAM Part III, which would have influenced students' song performances, and consequently the reliability of the test. However, the low test-retest reliability may also indicate the instability of young children's skill in singing songs. Possibly, the task of singing a song is not approached in the same manner twice even by the same child, and it may involve emotional and attitudinal components to a larger degree than single note or pattern echo singing tasks.

An unexpected result on the SAM Part III was that the posttest mean scores were lower than the pretest means for all classes, except E2 where the posttest mean increased slightly. This did not mean that all students' singing was less accurate after instruction, rather that the posttest was more difficult than the pretest. Both the posttest song (Here we go Looby Loo) and the pretest song (Farmer in the Dell) are in the same range and meter, and have four phrases and simple words. The melody of the posttest song has more melodic leaps, and it was thought that the melodic repetition in phrases 1, 2, and 3 would have balanced the song difficulty, but this apparently was not so. The first phrase of the pretest song is essentially one repeated note, and the more difficult melodic activity in the last two phrases was deemed to balance the simplicity of the first phrase. However, this was not the case. This discrepancy between the pretest-posttest level of difficulty necessitated that the results of each test be interpreted separately.

Although song singing accuracy may be difficult to measure and there were several problems with the test used to measure this skill, it must be assumed from the results that the experimental program did not improve the song singing accuracy of Grade 1 students to a significant extent. Possible reasons for the lack of improvement are the complexity of the song singing skill, and the short duration of the teaching program. The experimental classes' posttest means were higher than the control classes' posttest means, but not different enough for significance. In eight weeks students may not have been provided with sufficient opportunity to practice their singing skills. Most students were not working beyond Stage 5 in Gould's sequence (echoing longer patterns accurately with "loo") by the end of the instructional program. The instructional period was long enough to enable students to acquire or improve some basic singing skills (as shown by the significant improvement in single pitch and pattern singing), but may not have been long enough for them to apply these skills to the task of singing a song.

Problems with PMMA Results. Although there was a significant difference between the experimental and control groups' posttest scores on the Primary Measures of Music Audiation as analyzed by the analysis of covariance with pretest covariate, the repeated measures design showed no significant difference between pretest and posttest between experimental group and control group. The reason for the lack of significant difference in improvement may have been due to a ceiling effect. The mean score in Gordon's norm sample of 202 Grade 1 students was 29.8 (Gordon, 1979a, p. 87), and the pretest mean for the entire sample in this study was 31.46. The pretest scores may have been sufficiently high that there was not much "room" to gain.

Slightly more than ten percent of the students pattern-marked their PMMA test sheets, and consequently their scores were not indicative of their melodic perception and could not be used in the data analysis. There may have been characteristics which these eleven students had in common, if so, the deletion of their scores from the data may have affected the results.

Additional Results

The experimental instruction was effective in improving the singing accuracy of the experimental group as a whole.

Results of an additional t-test showed that experimental group students who were poor singers before instruction improved their singing significantly more than control group poor singers, indicating that the experimental instruction was beneficial for those who most needed improvement. A comparison of pretest and posttest scores showed that many experimental group poor singers developed to moderate or good singers.

There was a significant difference in singing gains between experimental group students and control group students with high melodic perception skills, indicating that the experimental instruction was also particularly effective for students with very good melodic perception skills before instruction. The melodic perception skills of these experimental students were utilized during singing activities in the experimental program, and assisted them in improving their vocal accuracy. Students with high melodic perception did not benefit greatly from control instruction; possibly the control activities did not challenge these students or enable them to use their capabilities.

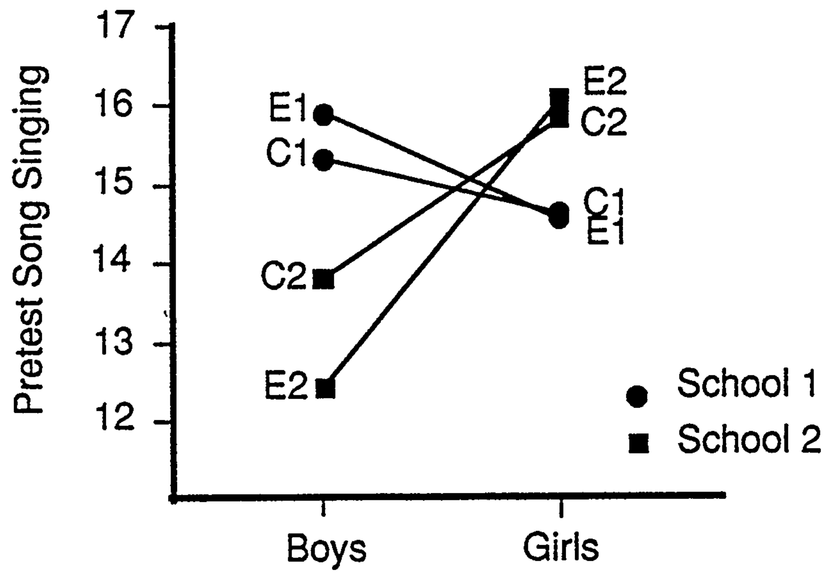
Experimental group students with low melodic perception made greater singing gains than control students with low melodic perception, but the difference was not significant. An examination of the pretest and posttest PMMA scores of experimental students in the low melodic perception group, revealed that the melodic perception of many of these students improved considerably as a result of instruction, although their average singing improvement was not as great as that of experimental students with high melodic perception.

These results suggest that improvement in melodic perception may be a necessary prerequisite to improvement in singing accuracy. The experimental instruction appears to have improved the melodic perception of students with poor melodic perception, and to have improved their singing accuracy, but not significantly more than the control groups' singing improvement. Students who began the experimental program with high melodic perception were able to significantly improve their singing.

The lack of a significant interaction between method of instruction and gender of students on the ANCOVA for SAM Part I and Part II indicates that the difference in singing accuracy after instruction between experimental group boys and girls was similar to the difference in singing accuracy after instruction between control group boys and girls. Boys and girls responded similarly to instruction.

A significant initial difference between schools was found for SAM Part III, and the pretest interaction between gender and class further elucidates this difference. In School 1, boys' mean scores in song singing were higher than girls' mean scores in both the experimental and control classes. In School 2, girls' mean scores were higher than boys', in both classes. Figure 9 presents the means for boys and girls by class for SAM Part III. Previous research indicated that girls tend to have higher singing accuracy scores. The reason for the boys' higher singing scores in School 1 may be the influence of the male principal at this school, who is reported to have sung and played guitar in student assemblies and at community gatherings. He possibly acted as a role model for singing for boys in School 1. The task of song singing appears to be affected by attitudinal influences to a larger degree than the single note echo singing task or the melodic pattern singing task.

Figure 9 SAM Part III Class Mean Scores by Gender



Secondary Comparisons

The pretest data of all students was also used to describe the relationship between singing accuracy and melodic perception, gender, and home musical environment in all students involved in the study before the influence of instruction.

The results of the present study concur with Petzold's findings (1963), that girls had higher mean scores in singing accuracy than boys, but that the difference was not significant. Considerable variation in individual achievement was found to exist within the overall group.

This study found a significant difference in the singing accuracy of students with high melodic perception skills, and students with low melodic perception skills. Both Bentley (1968) and Zwissler (1971) found significant differences between the melodic perception of accurate singers and inaccurate singers. Considered together, these results suggest that development in either one of these skills appears to be related to development in the other.

In the present study, students from home environments where singing and musical activities occurred frequently were found to have better singing, but not better melodic perception than students from homes with few musical influences. These results concur with Kirkpatrick (1962) and Brand (1986). Home environment seems to play a larger role in the development of singing accuracy skills than in the development of melodic perception skills. Singing is a more audible task than perceiving. A child in a home that does not encourage musical development may still learn to perceive melodic changes in music without instruction or encouragement. However, a child's attempts at singing will be obvious, and if not encouraged, the child may not have the much-needed opportunity to practice this skill, and may not have an attitude open to developing it in the future.

Recommendations for Further Research

Based on the results of this study, several recommendations are made.

1. In-depth research that examines the singing problems of individual students would be a valuable extension to the statistical results of this study. The particular procedures and exercises found to be helpful for students with specific singing problems, and sequences of vocal learning found for individual students could be established.

2. A similar study providing instruction over a longer period of time would allow for vocal instruction beyond the Stage 5 level to occur. With a longer

instructional period, an assessment could be made of the extent to which singing skills acquired through exercises would transfer to the singing of songs.

3. Further improvement on the singing test design and validation, including a comparison of test scores with singing observed and evaluated by experienced vocal educators, would add to the credibility of the numerical results. The design of tests for song singing should be seriously reviewed; test songs should be chosen and tested for level of difficulty in the way that previous studies have compared melodic patterns for difficulty level.

4. Research of an ethnographic or observational nature, dealing with the motivational and attitudinal aspects of learning to sing, might be able to answer some of the questions resulting from this study.

5. An examination of the influences that cause differential singing responses in boys and girls, and the testing of alternative educational approaches to encourage the singing development of boys would be a beneficial addition to the research in children's singing.

Although the present investigation dealt with only a limited number of children, the results have implications for music education in a broader field. In this study, Grade 1 children improved their singing accuracy when echo singing single pitches and when echo singing melodic phrases. That a significant improvement occurred in only eight weeks suggests that the Grade 1 students were ready for vocal instruction, and that the program was appropriate. Grade 1 children can improve their singing accuracy, particularly if they receive music instruction that is sequenced, beginning with vocal inflection exercises, then echo singing using neutral syllables in a narrow then gradually extended range. Activities to develop pitch concepts and pitch discrimination also appear to be important components, as well as individual singing and programming, and an emphasis on careful listening.

Although it is recommended that this study be replicated in other settings, and with older and younger children, the results are conclusive enough that the experimental program can be used by all music teachers who have accepted the responsibility for developing in-tune singing in their students.

Music can be such a powerful influence in an individual's life that opportunities for successful participation, including in-tune singing, should be provided for children from an early age. It is hoped that the results of this study will assist teachers to design appropriate lessons for promoting in-tune singing, and that the goal of joyful musical expression can be reached for all children.

BIBLIOGRAPHY

Apfelstadt, H. (l983). An investigation of the effects of melodic perception instruction on the pitch discrimination and vocal accuracy of kindergarten children. Dissertation Abstracts International, 44, 1719A. (University Microfilms No. 8316194)

Apfelstadt, H. (l984). The effects of melodic perception instruction on pitch discrimination and vocal accuracy of kindergarten children. Journal of Research in Music Education, 32, l5-24.

Atterbury, B. W. (l984a). Are you really teaching children how to sing? Music Educators Journal, 70(8), 43-45.

Atterbury, B. W. (l984b). Children's singing voices: a review of selected research. Bulletin of the Council for Research in Music Education, 80, 5l-63.

Bennett, P. (l986). A responsibility to young voices. Music Educators Journal, 73(l), 33-38.

Bentley, A. (l966). Musical ability in children and its measurement. New York: October House.

Bentley, A. (l968). Monotones. (Music Education Research Papers No. 1). London: Novello.

Best, J. W. (1986). Research in education (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Boardman, E. L. (l964). An investigation of the effect of preschool training on the development of vocal accuracy in young children. Dissertation Abstracts, 25, l245. (University Microfilms No. 64-8354)

Brand, M. (1986). Relationship between home musical environment and selected musical attributes of second-grade children. Journal of Research in Music Education, 34, 111-120.

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(1), 36-47.

Burroughs, G. E. R. (1971). Design and analysis in educational research. (Educational Monograph No. 8). Birmingham: University of Birmingham, School of Education.

Campbell, D. T. & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Choksy, L., & Brummit, D. (1987). 120 Singing games and dances for elementary schools. Englewood Cliffs, NJ: Prentice-Hall.

Cook, T. D., & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.

Davies, A. D. M. & Roberts, E. (l975). Poor pitch singing: a survey of its incidence in school children. Psychology of Music, 3(2), 24-36.

Dibble, C. A. (1983). Videotape pitch discrimination instruction of five-year-old children from different home musical environments. Dissertation Abstracts International, 44, 1369A.

Flowers, P., & 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.

Forcucci, S. L. (l975). Help for inaccurate singers. Music Educators Journal, 62(2), 57-61.

Gardner, H., Davidson, L., & McKernon, P. (1979). The acquisition of song: a developmental approach. In National Symposium on the Applications of Psychology to the Teaching and Learning of Music (pp. 301-317). Reston, VA: Music Educators National Conference.

Gay, L. R. (l98l). Educational research: Competencies for analysis and application (2nd ed.). Columbus, OH: Charles E. Merril.

Geringer, J. (1983). The relationship of pitch-matching and pitch-discrimination abilities of preschool and fourth-grade students. Journal of Research in Music Education, 31, 93-99.

Goetze, M. (1985). Factors affecting accuracy in children's singing. Dissertation Abstracts International, 46, 2955A. (University Microfilms No. 8528488)

Goetze, M., Cooper, N., & Brown, C. (1990). Recent research on singing in the general music classroom. Bulletin of the Council for Research in Music Education, 104, 16-37.

Gordon, E. (l97l). The psychology of music teaching. Englewood Cliffs, NJ: Prentice-Hall.

Gordon, E. (l977). Learning sequence and patterns in music. Chicago: G. I. A.

Gordon, E. (l979a). Primary measures of music audiation: Test manual. Chicago: G. I. A.

Gordon, E. (l979b). Developmental music aptitude as measured by the Primary Measures of Music Audiation. Psychology of Music, 7(1), 42-49.

Gordon, E. (l980). The assessment of music aptitude of very young children. Gifted Child Quarterly, 24, 107-111.

Gordon, E. (l986). A factor analysis of the Musical Aptitude Profile, the Primary Measures of Music Audiation and the Intermediate Measures of Music Audiation. Bulletin of the Council for Research in Music Education, 87, l7-25.

Gordon, E. (l986). The development of musicality from preschool through early childhood: Pedagogical implications. Paper presented at Music Educators National Conference, Anaheim, CA.

Gould, A. Oren. (l968a). Developing specialized programs for singing in the elementary school - final report. Washington, DC: U. S. Department of Health, Education and Welfare. (ERIC Document Reproduction Service No. ED 025 530).

Gould, A. Oren. (l968b). Finding and learning to use the singing voice: A manual for teachers. Washington, DC: U. S. Department of Health , Education and Welfare. (ERIC Document Reproduction Service No. ED 025 53l).

Gould, A. Oren. (l969). Developing specialized programs for singing. Bulletin of the Council for Research in Music Education, 17, 9-22.

Green, G. A. (1987). The effect of vocal modelling on pitch-matching accuracy of children in grades one through six. Dissertation Abstracts International, 48, 1410A.

Greenberg, M. (1979). Your children need music. Englewood Cliffs, NJ: Prentice-Hall.

Hattwick, M. (1933). The role of pitch level and pitch range in the singing of preschool, first grade, and second grade children. Child Development, 4, 281-291.

Jarjisian, C. (1981). The effects of pentatonic and/or diatonic pitch pattern instruction on the rote-singing achievement of young children. Disseration Abstracts International, 42, 2015A. (University Microfilms No. 8124581)

Jersild, A. & Bienstock, S. (l931). The influence of training on the vocal ability of three-year-old children. Child Development, 2, 272-291.

Jersild, A. & Bienstock, S. (l934). A study of the development of children's ability to sing. Journal of Educational Psychology, 25, 48l-503.

Jones, B. A. (1981). A comparative study of spatial reinforcement as a means for improving the pitch discrimination of seven year olds. Dissertation Abstracts International, 42, 592A. (University Microfilms No. 8117286)

Jones, M. (1979). Using a vertical-keyboard instrument with the uncertain singer. Journal of Research in Music Education, 27, 173-184.

Jordan-DeCarbo, J. (1982). Same/different discrimination techniques, readiness training, pattern treatment, and sex on aural discrimination and singing of tonal patterns by kindergartners. Journal of Research in Music Education, 30, 237-246.

Joyner, D. R. (1969). The monotone problem. Journal of Research in Music Education, 17, 115-124.

Kagan, J. (1978). The growth of the child: Reflections on human development. New York: W. W. Norton.

Kavanaugh, J. (1982). The development of vocal concepts in children: the methodologies recommended in designated elementary music series. Dissertation Abstracts International, 43, 2270A.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Kirkpatrick, W. C. (1962). Relationships between the singing ability of prekindergarten children and their home musical environment. Dissertation Abstracts, 23, 886.

Kramer, S. (1985). The effects of two different music programs on third and fourth grade children's ability to match pitches vocally. Dissertation Abstracts International, 46, 2609A. (University Microfilms No. 8524224)

Lehman, P. R. (1968). Tests and measurements in music. Englewood Cliffs, N. J.: Prentice-Hall.

Mehr, N. (1985). Helping children perceive melody. Music Educators Journal, 71(8), 29-31.

Moog, H. (1976). The development of musical experience in children of pre-school age. Psychology of Music, 4(2), 38-45.

Moore, D. L. (1973). A study of pitch and rhythm responses of five-year-old children in relation to their early musical experiences. Dissertation Abstracts International, 34, 6689A.

Nie, N. H., Hull, C. J., Jenkins, J. G., Steinbrenner, K., & Bent, D. J. (1983). Statistical package for the social sciences x. New York: McGraw-Hill.

Nunnally, J. (1972). Educational measurement and evaluation. New York: McGraw-Hill.

Patrick, L. R. (1978). An investigation of the pitch-matching abilities of first grade children. Dissertation Abstracts International, 40, 531A.

Petzold, R. (1963). The development of auditory perception of musical sounds by children in the first six grades. Journal of Research in Music Education, 11, 21-43.

Petzold, R. (1969). Auditory perception by children. Journal of Research in Music Education, 17, 82-87.

Phillips, K. (1984). Child voice training research: Song approach - Formalized training. Journal of Research in Singing, 8(1), 11-25.

Phillips, K. (1985). The effects of group breath-control on the singing ability of elementary students. Journal of Research in Music Education, 33, 179-191.

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. 76-19898)

Roberts, E., & Davies, A. (1975). Poor pitch singing: response of monotone singers to a program of remedial training. Journal of Research in Music Education, 23, 227-239.

Roberts, E., & Davies, A. (1976). A method of extending the vocal range of "monotone" school children. Psychology of Music, 4(1), 29-43.

Rosborough, K., Troncoso, L., & Piper, R. (1972). Teaching the young child to sing: A literature review with annotated bibliography. (Report No. TN-3-72-11). Los Alamitos, CA: Southwest Regional Laboratory for Educational Research and Development. (ERIC Document Reproduction Service No. ED 109 018).

Shelton, J. S. (1965). The influence of home musical environment upon musical response of first-grade children. Dissertation Abstracts, 26, 6765.

Shuter-Dyson, R. (1968). The psychology of musical ability. London: Methuen.

Sims, W., Moore, R., & Kuhn, T. (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, 10, 104-108.

Sinor, E., (1984). The singing of selected tonal patterns by preschool children. Dissertation Abstracts International, 45, 3299A. (University Microfilms No. 8501456)

Smale, M. (1987). An investigation of pitch accuracy of four- and five-year-old singers. Dissertation Abstracts International, 48, 2013A.

Small, A. R., & 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, 227-233.

Smith, R. (1963). The effect of group vocal training on the singing ability of nursery school children. Journal of Research in Music Education, 11, 137-141.

Steeves, C. (1984). The effect of Curwen-Kodály handsigns on pitch and interval discrimination within a Kodály curricular framework. Unpublished master's thesis, University of Calgary, Calgary, AB.

Stene, E. (1969). There are no monotones. Music Educators Journal, 55(8), 46-49, 117-121.

Welch, G. F. (1979a). Poor pitch singing: A review of the literature. Psychology of Music, 7(1), 50-58.

Welch, G. F. (1979b). Vocal range and poor pitch singing. Psychology of Music, 7(2), 13-31.

Welch, G. F. (1984). A schema theory of how children learn to sing in-tune: an empirical investigation. Proceedings, Stockholm Music Acoustics Conference 1983, 1, 323-332.

Welch, G. F. (1985). A schema theory of how children learn to sing in tune. Psychology of Music, 13(1), 3-18.

Welch, G. F. (1986a). A developmental view of children's singing. British Journal of Music Education, 3, 295-303.

Welch, G. F. (1986b). The potential for musical behavior in early childhood. Ohio Music Educators Association Journal, 13, 31-38.

Wolner, M., & Pyle, W. (1933). An experiment in individual training of pitch-deficient children. Journal of Educational Psychology, 24, 602-608.

Young, W. T. (1971). An investigation of the singing abilities of kindergarten and first grade children in east Texas. Washington, DC: U. S. Department of Health , Education and Welfare. (ERIC Document Reproduction Service No. ED 069 431).

Zimmerman, M. (1971). Musical Characteristics of Children. (From research to the music classroom No. 1). Washington, DC: Music Educators National Conference.

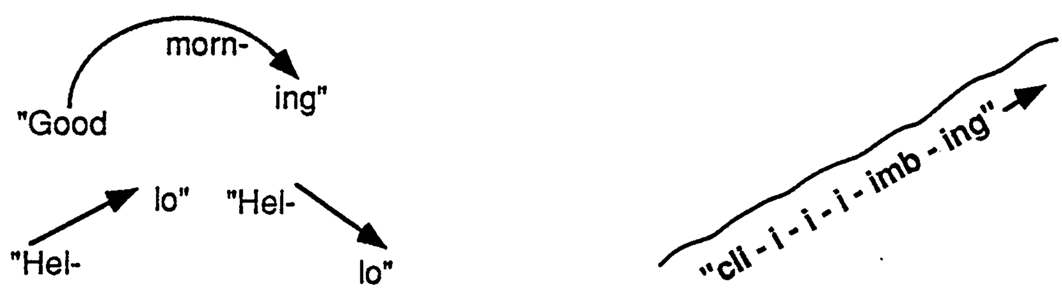
Zwissler, R. (1971). An investigation of the pitch discrimination skills of first grade children identified as "accurate singers" and those identified as "inaccurate singers". Dissertation Abstracts International, 32, 4057A. (University Microfilms No. 72-2947)

Appendix A

Gould's Speech to Song Sequence

Stage One - Speech Activities I

At the end of Stage One, the child should be able to control the pitch levels of his or her voice in speech and demonstrate an understanding of the concept that sounds can be pitched in higher and lower places. The first activities in Gould's sequence are designed to promote children's awareness that they can manipulate their spoken voice to create higher and lower sounds. This involves each child listening to his or her own voice and to the voices of others. Short phrases, greetings, questions and answers, and short poems will be spoken with exaggerated voice inflections. "Higher" and "lower" will be used to describe the sounds. Pitch contour will be shown with hands and with contour lines drawn on the board. The following are some examples of words with inflections shown:



An attempt should be made to motivate the children to become involved in the activities, and to develop in each child a sense of being special, that his or her efforts are valued. The teacher should establish the current skill level of each student, and monitor student progress throughout the instructional period.

Stage Two Speech Activities II

At the end of Stage Two, each child should be able to distinguish between his or her singing and speaking voice, and the speaking and singing voices of others. The activities at this stage are designed to develop the child's concept of the difference between the speaking voice and the singing voice, and encourage the child's skill in producing a singing quality with the voice. The children will echo the teacher's different voice qualities, identifying each. Some examples of the activities which will be used are:

Teacher: "This is my whispering voice." (children echo)

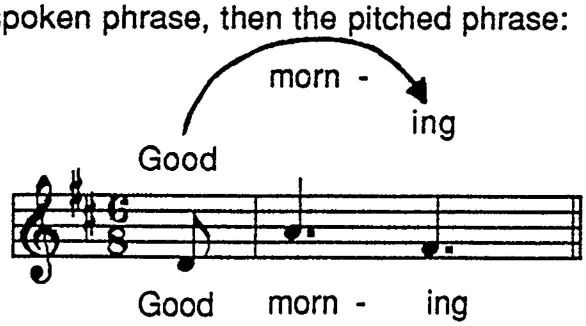
"This is my talking voice." (children echo)

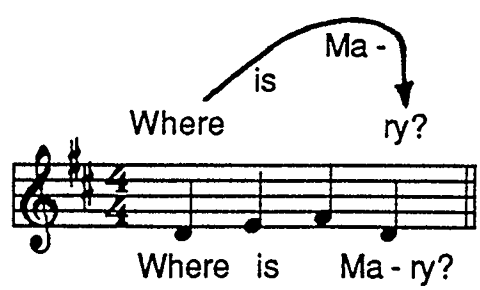
"This is my singing voice." (children echo)

Game-like activities will be used to ascertain whether the children can hear and produce different types of voices.

<u>Stage Three - Echo singing and Experiencing Unison in Song</u>

At the end of this stage, each child should be able to echo sing short (2, 3, or 4 notes) melodic patterns in his or her chosen range, which may be lower than the normal singing range for 5- to 7-year-olds, defined as D' to A' by Gordon (1971). In addition, each child should be able to identify when he or she is singing the same tones as someone else. Through the following activities, the researcher will attempt to lead children to control their voices to produce pitched intervals. Initially, the same phrases that were spoken in Stage One will be sung. These sung phrases should follow the same contour as the spoken phrases. The melodic contour of these short patterns will be shown with hand movements. The following examples provide the original spoken phrase, then the pitched phrase:





Through these techniques, the researcher will be attempting to develop the student's concept of sounds as being the same or different. The group and individuals should be able to identify when two people are "matching pitches", that is, when the singing sounds "the same". The child who cannot match the teacher's pitch will be allowed to continue to sing many patterns and songs in his or her comfortable range before attempting to move his or her voice (up, usually) to the range of children at that age who sing easily.

Stage Four - Finding the True Singing Voice

At the end of this stage, each child should be able to sing using an oo sound on two- or three-note echoed patterns in songs in a range closer to the desired singing range. The researcher will have the children echo the "oo" sound on short patterns from songs in an attempt to extend their singing range to the D' to A' range. Some Hallowe'en songs, train, owl, and wind songs contain such passages. Visual and verbal reinforcement of higher, lower or same pitch singing will continue to be used at this stage for the purpose of developing the child's awareness of his or her singing.

Stage Five - Expanding the Repertoire of Patterns Sung

At the end of this stage, before moving on to the next stage, each child should be able to sing the "oo" sound on longer patterns, then phrases. Slow legato singing will be encouraged as the child attempts to accurately sing longer patterns and phrases, so that he or she will have time to hear the intervals and coordinate the mental sound images formed in his or her mind with the motor responses made with the voice. The child should be able to demonstrate considerable proficiency at matching extended patterns, phrases and songs using the oo sound in a singing range, before moving on to the next stage.

Stage Six - Type A Songs

At the end of Stage Six, each child should be able to sing with words, songs that begin on the tonic and include descending scale passages, as well as demonstrate an understanding of "ascending and descending melodic line".

Stage Seven - Type B Songs

At the end of this stage, each child should be able to sing songs that begin on the mediant or the dominant note, as well as the tonic.

Stage Eight - Type C Songs

At the end of this stage, each child should be able to sing with words, songs that begin on a higher tone than the low tonic and have one large skip to an upper tonic, no higher than D.

During the last three stages, after basic skills in tone matching have been demonstrated, the child's singing repertoire will be expanded and consolidated to include the melodic patterns described. Melodic contour will continue to be indicated by teacher and student using chalkboard diagrams and hand and arm motions, to create visual associations with the pitches of songs and patterns. All students may not reach Stage Eight during the course of the experiment. When responding individually, students will be taught at the stage at which they are functioning.

Appendix B

SINGING ACHIEVEMENT MEASURES -- Final Version

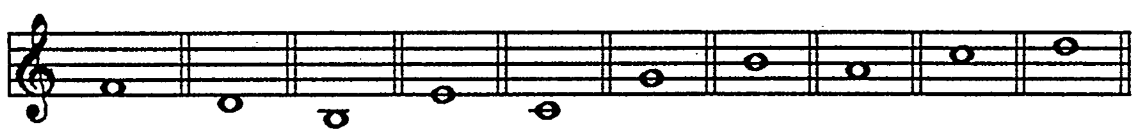
All test items will be sung by a soprano voice, tape-recorded using a Technics tape deck and Sony F-V30T microphone. Students' responses will also be tape-recorded. Identical instructions will be spoken to each student by the researcher before starting the test tape.

INSTRUCTIONS

These instructions are to be spoken to each student before he or she hears the test items.

Part I - Single Pitch Singing

"I am going to start the tape. On the tape you will hear a woman singing. Try to sing back or echo the same sound that you hear." (10 pitches, each repeated on the tape and sung twice, will be presented).



Part II - Pattern Singing

"Now you will hear little songs without words. Try to sing back the same sounds that you hear on the tape." (10 patterns, each repeated on the tape to be sung twice by the student, will be presented).

Part III - Free Song

"Let's sing The Farmer in the Dell." The investigator sings the first two words ("The farmer ---") on the pitches C' F' F' and encourages the child to begin singing the song.

The posttest song will be "Here we go Looby Loo" and the directions will remain the same.

RATING SCALE FOR SCORING - SINGING ACHIEVEMENT MEASURES

Part I Scoring

Objective: The student will correctly sing one pitch after it is sung by the voice on the tape.

Score one point for each pitch sung correctly. A short slide into the correct pitch scores one point. Each pitch will be presented twice; score each attempt. (20 points maximum)

Part II Scoring

Objective: The student will correctly sing the pattern, maintaining pitch level (i.e. not transposing the pattern). Scoring places a greater value on the maintenance of pitch level over the maintenance of pattern. The purpose is to evaluate singing in the desired pitch range (D' to A').

Evaluate each pattern according to the following scale:

The PITCH LEVEL is maintained and -

5 points: - all pitches are correct.

4 points: - 1 or 2 small inaccuracies in interval size, although interval direction is correct.

3 points: - several small or 1 or 2 large inaccuracies in interval size, interval direction is the same.

2 points: - several small or 1 or 2 large inaccuracies in interval size, interval direction is not entirely the same as the original.

The PITCH LEVEL is not maintained and -

3 points: - interval size and interval direction are the same as original, i.e. the pattern is correctly transposed.

2 points: - 1 or 2 small inaccuracies in interval size, interval direction is the same.

1 point: - great inaccuracies in interval size, interval direction is not entirely the same as the original.

0 points: - response is not in a singing voice.

If the child does not sing the entire pattern, rate what he or she has sung, and deduct one point. If a repeated note is omitted, or the rhythm changed, a point will not be subtracted. Each of the ten patterns will be presented twice, sung twice by the student, and scored separately. Maximum score for twenty (10 x 2) correct patterns is 100 points.

Part III Scoring -- Free Song

Score 5 points for each phrase of The Farmer in the Dell or Here we go Looby Loo sung correctly. Use the same scoring scale as used in Part II (20 points maximum).

Total Possible Points:

Part I -	Single Pitches	20
Part II -	Patterns	100
Part III -	Free Song	20
		140 points

Appendix C

Home Musical Environment Questionnaire

Scoring Scheme for the Home Musical Environment Questionnaire

Convert responses to numerical scores: if "1" is circled, score 6 points, if "2" is circled, score 5 points, etc.

"Total"

In Part I, score questions 1, 2, 3, 4, 5, 11, 12, 13; in question 13, add scores for classical/semi-classical and opera/musicals only. The maximum score for each question is 6 points, except for question 13, where a maximum of 12 is possible. In Part II, score questions 1, 2, 3, 4, 5, 6, and evaluate questions 7, 8, and 9 together for a maximum of 6 points. The maximum score for each question is 6 points, except for question 5, where a maximum of 12 is possible.

"Siblings"

Score questions 7, 8, 9, and 10 in Part I. The maximum score for each question is 6 points.

"Parents"

Score questions 2 and 5 in Part I. The maximum score for each question is 6 points.

"Child"

Score question 1 in Part II only. The maximum score is 6 points.

Questionnaire for Parents/Guardians

Name of Child:______________________________________ Birthdate:______________
Sex:____

Dear Parents/Guardians,

Your responses to the following questions will assist the researcher in determining the musical environment of your child. All information will be kept confidential. Please answer as accurately as possible.

Unless otherwise indicated, circle the number beside each question that corresponds to the response below which best applies to your situation. Please answer every question.

1. daily	2. at least weekly	3. at least monthly	4. at least twice a year	5. at least yearly	6. never

Part I

1. Do parents/guardians listen to music in the home? 1 2 3 4 5 6

2. Do parents/guardians sing at home? 1 2 3 4 5 6

3. Do parents/guardians play musical instruments at home? 1 2 3 4 5 6
   If so, identify, instruments: ____________________________

4. Do parents/guardians currently participate in any musical groups?1 2 3 4 5 6
   If so, identify type: __________________________________

5. Do parents/guardians help the child to learn songs? 1 2 3 4 5 6

6. Do family members take the child to concerts? 1 2 3 4 5 6

7. How many brothers does the child have? ____ older; ____ younger
   How many sisters does the child have? ____ older; ____ younger

8. Do brothers or sisters listen to music? 1 2 3 4 5 6

9. Do brothers or sisters sing at home? 1 2 3 4 5 6

10. Do brothers or sisters play any musical instruments at home? 1 2 3 4 5 6
    If so, identify instruments: ____________________________

11. Do brothers or sisters participate in any musical groups in or out of school?
    1 2 3 4 5 6
    If so, identify type: __________________________________

12. Do brothers or sisters help the child to learn songs? 1 2 3 4 5 6

13. How often is -

big band music heard in your home?	1	2	3	4	5	6
classical or semi-classical?	1	2	3	4	5	6
country and western?	1	2	3	4	5	6
folk?	1	2	3	4	5	6
rock?	1	2	3	4	5	6
opera or musicals?	1	2	3	4	5	6

1. daily	2. at least weekly	3. at least monthly	4. at least twice a year	5. at least yearly	6. never

Part II

1. Does the child sing spontaneously while playing with toys, games etc.? 1 2 3 4 5 6

2. Does the child sing with -
   adults? 1 2 3 4 5 6
   brothers or sisters? 1 2 3 4 5 6
   TV/radio/records/tapes? 1 2 3 4 5 6

3. Does the child play any musical instruments in the home? 1 2 3 4 5 6
   If so, identify instruments: ____________________

4. Does the child listen to music (radio,records, tapes) in the home?1 2 3 4 5 6

5. How often does the child listen to -
   children's songs in or out of the home? 1 2 3 4 5 6
   classical or semi-classical? 1 2 3 4 5 6
   other (identify type) ____________________ 1 2 3 4 5 6

6. Has the child participated in any organized music activities or classes outside school? 1 2 3 4 5 6
   If so, indicate the specific nature of the classes:
   ____
   ____

7. Did the child attend preschool, Early Childhood Services (E. C. S.), or kindergarten?
   ___ yes; ___ no

   If so, for what length of time? (circle one):
   ___ 2 years or more; ___ more than one year; ___ full year; ___ less than one year.

   If so, what kinds of musical activities did the child experience in that program?
   ___ singing ___ moving to music ___ listening
   ___ playing instruments ___ not sure

8. Please describe any <u>other</u> musical activities participated in by members of your family:
   ____
   ____
   ____

9. Please describe any <u>other</u> musical activities participated in by your child:
   ____
   ____
   ____

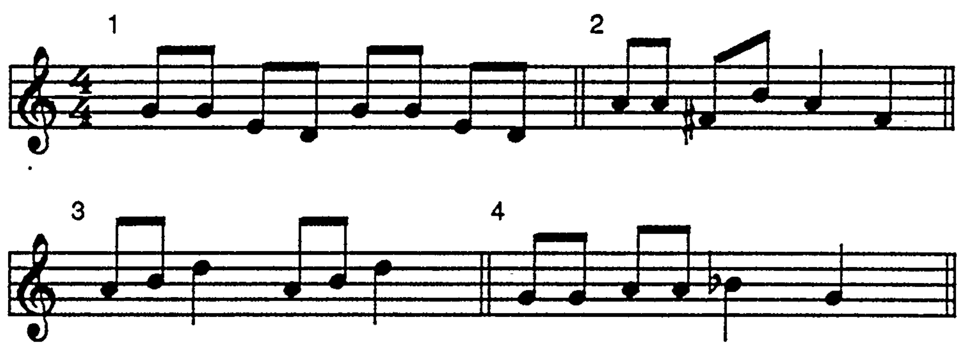
Appendix D

Preliminary Versions of *Singing Achievement Measures* Part II

Test Version1

Test Version 2

This version was tested with daycare children.



Comparison of Daycare and Sinor Pattern Rankings

The patterns in version 2 of the test were ranked according to the difficulty the daycare children had in singing the pattern accurately. The number of children who scored 4 or 5 determined the ranking of the pattern. For ease of comparison, the patterns are transposed to the key of G major or E minor. Patterns were ordered so that an lower number designated an easier pattern.

Sinor (1984) tested 48 four-note patterns with 96 children and ranked the patterns according to level of difficulty.

#7, Rank 1 (SAM #2) #27, Rank 1

#6, Rank 6 (SAM #7) #27, Rank 1

#4, Rank 3 (SAM #6) #3, Rank 2

#10, Rank 7 (SAM #9) #44, Rank 3

#2, Rank 4 (SAM #1) #40, Rank 7

#8, Rank 9 (SAM #4) #13, Rank 8

#5, Rank 8 (omitted from SAM) #4, Rank 14

#3, Rank 5 (SAM #8) #1, Rank 19

#9, Rank 10 (SAM #10) #16, Rank 35

#1, Rank 2 (SAM #3) #9, Rank 46

In the final version of the test, the order of the patterns was changed, based on daycare children's responses and the pattern difficulty found by Sinor. Pattern #5 was replaced with: 

See Appendix B for the final version of Singing Achievement Measures Part II.

Appendix E

[Pilot Test]

Information Letter for Parents/Guardians

Dear Parents/Guardians,

This letter is to inform you that your child, with your consent, will be participating in a research study at school in the 1987-88 school year. The study concerns children's singing, and the purpose of this research is to assist music teachers in developing more effective music programs at the Grade One level.

All the Grade One students participating will be given a singing test and a written aural pitch perception test; after one week, the tests will be re-administered. Both these tests were designed to be used with Grade One students. The results of the tests will be kept anonymous and will not be used for grading purposes. At no time will your child be identified in connection with any of the results of the study.

The project is being conducted as part of the research required for the degree of Master of Music and is under the supervision of Lois Choksy, Professor of Music at the University of Calgary.

If you are willing to have your child participate, please fill out the enclosed Consent Form and return it to the classroom teacher within the next few days. The success of this project depends on your support since the number of students involved is limited. If you have any further questions, please call me at home (239-1876). Your cooperation and support are greatly appreciated.

Yours sincerely,

Elizabeth Fleming
Graduate Student
The University of Calgary

[E1, E2]

Information Letter for Parents/Guardians

Dear Parents/Guardians,

This letter is to inform you that your child, with your consent, will be participating in a research study at school in the 1987-88 school year. The project is designed to determine the effects of a specific program of music instruction on children's singing ability.

The Grade One students will receive three half-hour music classes per week for eight weeks as their regular music programme. All the students participating will be given a written aural pitch perception test and a singing test immediately before and after the eight week period. Both these tests were designed to be used with Grade One students. The results of the tests will be kept anonymous and will not be used for grading purposes. Children will be graded for report card purposes on their progress, attitude and participation in music classes. At no time will your child be identified in connection with any of the results of the study.

There are no risks involved in this type of study. Your child will develop the basic musical skills of a Grade One music program under the instruction of a graduate student in music at the University of Calgary. The study is being conducted as part of the research required for the degree of Master of Music and is under the supervision of Lois Choksy, Professor of Music at the University of Calgary.

If you are willing to have your child participate, please fill out the enclosed Consent Form and Questionnaire and return them to the classroom teacher within the next few days. The success of this project depends on your support since the number of students involved is limited. If you have any further questions, please call me at home (239-1876). Your cooperation and support are greatly appreciated.

Yours sincerely,

Elizabeth Fleming
Graduate Student
The University of Calgary

[C1, C2]

Information Letter for Parents/Guardians

Dear Parents/Guardians,

This letter is to inform you that your child, with your consent, will be participating in a research study at school in the 1987-88 school year. The study concerns children's singing, and an assessment will be made of each participant's singing ability at the beginning and end of an eight week period.

The Grade One students will receive three half-hour music classes per week for eight weeks as their regular music programme. All the students participating will be given a written aural pitch perception test and a singing test immediately before and after the eight week period. Both these tests were designed to be used with Grade One students. The results of the tests will be kept anonymous and will not be used for grading purposes. Children will be graded for report card purposes on their progress, attitude and participation in music classes. At no time will your child be identified in connection with any of the results of the study.

There are no risks involved in this type of study. Your child will develop the basic musical skills of a Grade One music program under the instruction of a graduate student in music at the University of Calgary. The study is being conducted as part of the research required for the degree of Master of Music and is under the supervision of Lois Choksy, Professor of Music at the University of Calgary.

If you are willing to have your child participate, please fill out the enclosed Consent Form and Questionnaire and return them to the classroom teacher within the next few days. The success of this project depends on your support since the number of students involved is limited. If you have any further questions, please call me at home (239-1876). Your cooperation and support are greatly appreciated.

Yours sincerely,

Elizabeth Fleming
Graduate Student
The University of Calgary

Consent Form for Parents/Guardians

I, the undersigned, consent to have my child, ____________________ participate in the research study "The effect of vocal instruction using Gould's sequence on the melodic accuracy of Grade One students."

I am aware of the aims and methods of the research and the nature of student involvement.

At any time during the experimental period, I understand that I may withdraw this consent and that the investigator has the corresponding right to terminate the student's involvement.

Date ________________ Signature ________________________________

Teacher Consent Form

I, the undersigned, consent to have the Grade One class I teach at ______________________________________(name of school) participate in the research study "The effect of vocal instruction using Gould's sequence on the melodic singing accuracy of Grade One students."

I am aware of the aims and methods of the research, and the nature of student involvement.

At any time during the experimental period, I understand that I may withdraw this consent and that the investigator has the corresponding right to terminate the student's involvement.

Date ________________ Signature ________________________________

Principal Consent Form

I, the undersigned, consent to have the Grade One classes at________________________________(name of school) participate in the research study "The effect of vocal instruction using Gould's sequence on the melodic singing accuracy of Grade One students."

I am aware of the aims and methods of the research, and the nature of student involvement.

At any time during the experimental period, I understand that I may withdraw this consent and that the investigator has the corresponding right to terminate the student's involvement.

Date _______________ Signature _________________________________

Appendix F

Sample Lesson Plans

Experimental Class 1 and Experimental Class 2

Thursday November 12

Greeting - "Good morning boys and girls", (A' F#' B' A' A' F#')

Warm-up - "Who is wearing blue today?" (A' A' F#' B' A' A' F#')

Speech/Vocal Skills - Choose students who are singing at Stage Four level for the first three exercises. Show higher and lower sounds with arm motions. Encourage the class to listen and ask them whether the child's singing was in the same place as the teacher's. After some students, ask the class to echo the individual's singing.

- "The wind goes oo ------ oo, oo ------- oo". Echo sing oo on A' - F#'.
- "The owl goes 'oo' 'oo' 'oo' 'oo' ", sung on A'.
- Train sounds: "choo, choo", sung on A', B', C'', or D''.

Choose students who are singing at Stage Five level for the following exercises.

- "Fuzzy wuzzy was a bear" (A' F#' A' F#' D' D' D') - have the group, then individuals echo this phrase using "loo".
- "The Muffin Man" - Have students identify song from rhythm and notes without words. Individuals sing "yes I know the muffin man" on D' G' G' A' B' G' G' using the syllable "loo", or with words if they have been singing accurately.
- "Sing a song like a small flute blowing" (E' E' A' A' A' G' G' A' E') - have the group, then individuals echo this phrase using "loo", or words.

Familiar Song - "Doggie Doggie" encourage listening and accurate singing.

Game - "Doggie Doggie", guessing game with individual singing.

Beat and Rhythm Skills - "Hey Betty Martin", sing and tap beat, have students suggest body places to tap the beat. "Bell Horses", sing and tap beat, sing and march the beat.

Listening - "Land of the Silverbirch", children listen to teacher's singing and answer questions about the words.

Control Class 1 and Control Class 2

Thursday November 12

Greeting - "Good morning boys and girls", (A' F#' B' A' A' F#')

New and Familiar Songs - Sing, add actions where suitable:

"Hot Cross Buns"

"Peas Porridge Hot"

"Ten in a Bed"

"Rover"

"Three Crows"

"Little Peter Rabbit"

"In a Cabin by a Wood"

"Jim Along Josie"

Game - "Doggie Doggie", guessing game with individual singing.

Beat and Rhythm Skills - "Hey Betty Martin", sing and tap beat, have students suggest body places to tap the beat. "Bell Horses", sing and tap beat, sing and march the beat.

Listening - "Land of the Silverbirch", children listen to teacher's singing and answer questions about the words.

Appendix G

Observation Form

Classes ______________ Date _______________ Time _____________
Name of Observer ___________________________

1. Did you observe any of the following activities in the experimental group lesson? (place a "checkmark" each time the activity is observed)

   - exercises in speech inflection: ________________
   - exercises in differentiating between whispering, talking, and singing voice: ________________
   - individuals and the group attempt to match short phrases sung by the teacher or another student: ________________
   - students identify sounds as being the "same": ________________
   - students identify and demonstrate with hand and arm motions, "higher" and "lower" sounds: ________________
   - exercises involving the oo sound (individual and group responses): ________________
   - short phrases of songs sung by individuals in their comfortable range (lower range): ________________

2. Estimate the approximate total amount of lesson time spent involved in the above activities in the experimental group lesson: __________ min.

3. Were any of the above vocal technique activities included in the control group lesson? (circle yes or no) yes no
   If yes, identify which activity: ______________________________

4. Was there an equal amount of time spent -
   in the experimental group lesson on vocal technique activities
   and
   in the control group lesson singing additional songs and reinforcing non-melodic concepts (rhythmic, timbral)?
   Circle one: yes no

   If no, was more time spent involved in the above activities in the experimental group or in the control group? ______________________

5. Apart from the time spent in the activities outlined in question #1, was the remainder of the lesson taught in the same way to both experimental and control groups?
   Circle: yes no
   If no, identify any inconsistencies you noticed:

____

____

Appendix H

Scorers' Ratings and Student Performance Scores for *Singing Achievement Measures*

PRETEST *Singing Achievement Measures*

Experimental Class 1 - Pretest (n = 23)

		Scorers' Ratings									Performance Scores		
		Scorer 1 SAM Part			Scorer 2 SAM Part			Scorer 3 SAM Part					
Student Subject Code	Sex	I	II	III	I	II	III	I	II	III	SAM Part I	SAM Part II	SAM Part III
74	F	8	70	16	7	72	13	9	68	16	8.00	70.00	15.00
86	M	0	26	13	0	27	12	0	21	11	0	24.67	12.00
78	F	6	39	13	5	38	12	5	39	14	5.33	38.67	13.00
101	M	12	67	19	9	59	15	10	58	16	10.33	61.33	16.67
6	M	20	97	20	20	94	18	20	91	19	20.00	94.00	19.00
83	F	10	52	12	7	49	12	9	57	12	8.67	52.67	12.00
100	F	11	36	17	9	41	13	10	39	15	10.00	38.67	15.00
66	M	20	98	20	20	85	16	20	89	18	20.00	90.67	18.00
45	M	14	48	17	9	50	16	12	46	13	11.67	48.00	15.33
50	F	18	45	15	15	48	16	15	53	16	16.00	48.67	15.67
71	M	6	41	16	5	36	13	4	41	14	5.00	39.33	14.33
39	F	20	74	19	18	65	19	17	63	17	18.33	67.33	18.33
8	M	5	59	15	3	58	15	3	54	16	3.67	57.00	15.33
48	F	20	62	16	20	67	15	20	64	16	20.00	64.33	15.67
76	M	7	89	15	6	78	13	6	74	14	6.33	80.33	14.00
46	M	20	65	14	15	62	16	18	73	15	17.67	66.67	15.00
69	M	0	39	16	0	41	12	0	28	11	0	36.00	13.00
75	M	5	28	11	1	36	11	4	28	12	3.33	30.67	11.33
60	M	8	82	20	7	77	20	7	74	18	7.33	77.67	19.33
52	F	6	28	11	3	32	13	5	32	12	4.67	30.67	12.00
18	M	19	80	20	18	75	20	19	68	20	18.67	74.33	20.00
11	M	19	93	20	18	89	19	18	92	19	18.33	91.33	19.33
1	M	15	69	16	16	66	16	16	67	16	15.67	67.33	16.00

Experimental Class 2 - Pretest (n = 25)

Student Subject Code	Sex	Scorers' Ratings: Scorer 1 SAM Part I	Scorer 1 SAM Part II	Scorer 1 SAM Part III	Scorer 2 SAM Part I	Scorer 2 SAM Part II	Scorer 2 SAM Part III	Scorer 3 SAM Part I	Scorer 3 SAM Part II	Scorer 3 SAM Part III	Performance Scores: SAM Part I	SAM Part II	SAM Part III
23	F	19	71	16	18	69	12	20	74	14	19.00	71.33	14.00
90	M	4	40	10	4	47	10	4	46	8	4.00	44.33	9.33
73	F	6	44	14	5	31	12	7	32	13	6.00	35.67	13.00
85	M	7	49	18	4	44	17	4	46	17	5.00	46.33	17.33
87	F	5	46	13	5	52	14	8	47	16	6.00	48.33	14.33
4	M	3	46	14	3	40	11	3	41	12	3.00	42.33	12.33
2	M	10	42	13	7	45	13	9	44	12	8.67	43.67	12.67
42	F	16	62	16	12	58	15	14	60	14	14.00	60.00	15.00
80	M	20	74	20	20	74	19	20	66	19	20.00	71.33	19.33
61	M	4	33	13	3	39	7	2	32	8	3.00	34.67	9.33
59	M	10	62	12	9	66	12	12	65	12	10.33	64.33	12.00
13	M	11	59	13	11	48	10	13	51	12	11.67	52.67	11.67
3	M	15	47	17	9	48	15	12	52	15	12.00	49.00	15.67
44	F	20	89	18	20	90	19	20	86	19	20.00	88.33	18.67
94	F	18	70	20	18	72	19	18	73	20	18.00	71.67	19.67
7	F	7	50	15	3	47	14	7	44	15	5.67	47.00	14.67
31	M	17	60	12	16	64	14	18	67	15	17.00	63.67	13.67
37	M	18	37	15	17	38	14	20	44	13	18.33	39.67	14.00
97	F	20	92	20	19	89	19	20	86	19	19.67	89.00	19.33
9	M	4	22	13	4	30	10	4	32	11	4.00	28.00	11.33
33	M	6	17	8	4	27	6	6	26	6	5.33	23.33	6.67
51	M	14	60	12	12	58	12	15	58	13	13.67	58.67	12.33
10	M	3	42	11	2	47	9	3	49	7	2.67	46.00	9.00
104	M	18	72	14	16	71	9	16	69	10	16.67	70.67	11.00
53	M	9	41	16	5	49	12	8	54	13	7.33	48.00	13.67

Control Class 1 - Pretest (n = 27)

Student Subject Code	Sex	Scorer 1 SAM Part I	II	III	Scorer 2 SAM Part I	II	III	Scorer 3 SAM Part I	II	III	Performance Scores SAM Part I	SAM Part II	SAM Part III
64	F	14	51	13	11	45	11	13	40	13	12.67	45.33	12.33
35	F	17	52	13	16	60	19	17	64	18	16.67	58.67	16.67
102	M	15	57	17	14	49	13	14	56	13	14.33	54.00	14.33
25	M	19	93	18	20	89	19	19	92	20	19.33	91.33	19.00
91	F	20	80	17	20	78	15	20	78	15	20.00	78.67	15.67
98	M	15	53	15	14	50	14	15	54	13	14.67	52.33	14.00
12	M	19	63	16	18	76	15	19	75	17	18.67	71.33	16.00
82	M	6	65	17	3	61	15	4	67	17	4.33	64.33	16.33
56	M	11	45	11	7	41	11	11	45	10	9.67	43.67	10.67
43	M	6	25	19	4	47	13	6	50	16	5.33	40.67	16.00
27	F	17	57	14	17	63	13	17	67	13	17.00	62.33	13.33
17	F	16	65	15	16	76	16	17	72	16	16.33	71.00	15.67
79	M	7	44	14	6	34	10	7	38	8	6.67	38.67	10.67
81	M	14	71	18	12	71	16	12	67	15	12.67	69.67	16.33
93	F	5	38	17	4	48	16	4	51	17	4.33	45.67	16.67
58	F	3	53	13	1	44	15	3	46	13	2.33	46.67	13.67
20	F	15	68	17	12	67	11	14	66	16	13.67	67.00	14.67
77	M	11	61	14	10	52	13	11	51	14	10.67	54.67	13.67
84	F	0	45	17	0	39	14	0	38	18	0	40.67	16.33
26	M	17	90	20	17	83	18	17	82	20	17.00	85.00	19.33
62	M	20	88	18	20	91	19	20	84	19	20.00	87.67	18.67
14	M	0	47	17	0	55	14	0	47	12	0	49.67	14.33
41	M	12	62	14	4	55	11	8	53	14	8.00	56.67	13.00
70	M	17	48	17	15	47	14	15	41	15	15.67	45.33	15.33
36	F	18	56	11	12	53	11	15	51	12	15.00	53.33	11.33
16	M	18	45	15	17	54	15	17	60	13	17.33	53.00	14.33
47	M	20	57	19	20	54	18	20	58	16	20.00	56.33	17.67

Control Class 2 - Pretest (n = 25)

Student Subject Code	Sex	Scorers' Ratings									Performance Scores		
		Scorer 1 SAM Part			Scorer 2 SAM Part			Scorer 3 SAM Part			SAM Part I	SAM Part II	SAM Part III
		I	II	III	I	II	III	I	II	III			
24	M	10	75	19	10	79	17	10	72	19	10.00	75.33	18.33
19	M	19	84	20	19	78	17	19	76	16	19.00	79.33	17.67
38	M	2	47	14	2	44	13	2	42	13	2.00	44.33	13.33
15	M	2	30	15	1	34	15	2	31	17	1.67	31.67	15.67
55	M	18	83	20	16	78	19	19	73	19	17.67	78.00	19.33
49	M	18	50	19	17	58	19	19	59	18	18.00	55.67	18.67
5	M	16	53	14	16	60	14	17	56	13	16.33	56.33	13.67
21	F	5	27	11	2	26	11	4	31	11	3.67	28.00	11.00
57	M	9	42	11	9	38	12	9	39	11	9.00	39.67	11.33
29	M	17	67	14	11	63	14	13	61	14	13.67	63.67	14.00
67	F	19	88	19	18	78	19	19	77	19	18.67	81.00	19.00
96	F	20	100	20	20	98	19	20	97	19	20.00	98.33	19.33
99	M	18	38	18	19	45	17	18	47	19	18.33	43.33	18.00
40	M	6	39	12	6	37	10	6	35	13	6.00	37.00	11.67
32	F	20	96	19	20	94	20	20	97	20	20.00	95.67	19.67
34	M	2	37	9	0	29	7	2	23	3	1.33	29.67	6.33
65	F	20	58	15	15	53	13	19	59	13	18.00	56.67	13.67
95	M	3	0	0	3	0	0	5	0	8	3.67	0	2.67
105	M	11	66	15	8	60	13	9	58	14	9.33	61.33	14.00
103	M	2	45	17	2	48	16	3	49	16	2.33	47.33	16.33
63	M	7	34	14	5	32	10	7	31	10	6.33	32.33	11.33
54	M	1	33	14	0	37	11	0	35	14	.33	35.00	13.00
28	M	5	33	14	3	33	15	4	34	15	4.00	33.33	14.67
92	F	2	21	14	2	28	9	4	28	13	2.67	25.67	12.00
89	M	6	37	13	5	33	10	4	30	13	5.00	33.33	12.00

POSTTEST *Singing Achievement Measures*

Experimental Class 1 - Posttest (n = 23)

		Scorers' Ratings									Performance Scores		
Student Subject Code	Sex	Scorer 1 SAM Part I	II	III	Scorer 2 SAM Part I	II	III	Scorer 3 SAM Part I	II	III	SAM Part I	SAM Part II	SAM Part III
74	F	6	89	20	6	87	17	6	87	17	6.00	87.67	18.00
86	M	4	54	12	2	49	11	3	45	11	3.00	49.33	11.33
78	F	8	62	12	6	54	12	6	60	12	6.67	58.67	12.00
101	M	18	82	16	15	73	14	19	80	15	17.33	78.33	15.00
6	M	20	99	19	20	96	18	20	92	19	20.00	95.67	18.67
83	F	19	62	15	17	67	10	19	64	10	18.33	64.33	11.67
100	F	20	80	14	18	65	12	18	68	11	18.67	71.00	12.33
66	M	20	98	19	19	87	16	20	95	18	19.67	93.33	17.67
45	M	06	48	12	2	56	12	6	55	11	4.67	53.00	11.67
50	F	20	64	15	16	59	11	17	57	14	17.67	60.00	13.33
71	M	16	75	19	13	62	16	14	64	17	14.33	67.00	17.33
39	F	20	93	19	19	84	16	19	82	16	19.33	86.33	17.00
8	M	18	88	19	19	83	17	19	81	18	18.67	84.00	18.00
48	F	20	79	18	19	76	16	20	74	19	19.67	76.33	17.67
76	M	20	98	15	20	92	13	19	84	14	19.67	91.33	14.00
46	M	20	79	16	19	77	13	20	78	14	19.67	78.00	14.33
69	M	4	48	12	4	45	11	5	44	12	4.33	45.67	11.67
75	M	11	63	12	7	56	9	8	53	10	8.67	57.33	10.33
60	M	20	97	20	20	89	18	20	88	18	20.00	91.33	18.67
52	F	13	35	11	7	36	8	11	36	7	10.33	35.67	8.67
18	M	20	92	20	19	84	16	20	87	19	19.67	87.67	18.33
11	M	20	96	18	19	91	18	18	92	19	19.00	93.00	18.33
1	M	20	91	12	19	64	12	20	65	15	19.67	66.67	13.00

Experimental Class 2 - Posttest (n = 25)

Student Subject Code	Sex	Scorers' Ratings: Scorer 1 SAM Part I	II	III	Scorer 2 SAM Part I	II	III	Scorer 3 SAM Part I	II	III	Performance Scores: SAM Part I	SAM Part II	SAM Part III
23	F	20	92	17	20	79	14	20	83	17	20.00	84.67	16.00
90	M	11	57	13	8	54	10	11	61	12	10.00	57.33	11.67
73	F	6	81	15	6	71	12	7	70	14	6.33	74.00	13.67
85	M	20	96	19	20	85	13	20	81	17	20.00	87.33	16.33
87	F	11	59	16	10	54	11	12	59	14	11.00	57.33	13.67
4	M	11	52	14	7	52	11	11	55	11	9.67	53.00	12.00
2	M	19	79	15	18	65	10	19	73	11	18.67	72.33	12.00
42	F	20	85	16	18	68	11	20	76	14	19.33	76.33	13.67
80	M	20	97	19	20	86	17	20	90	18	20.00	91.00	18.00
61	M	11	58	16	8	52	12	11	49	11	10.00	53.00	13.00
59	M	19	66	17	17	73	14	19	74	12	18.33	71.00	14.33
13	M	20	75	18	20	69	15	19	70	18	19.67	71.33	17.00
3	M	15	49	14	15	48	13	17	55	13	15.67	50.67	13.33
44	F	20	93	19	19	94	19	20	94	19	19.67	93.67	19.00
94	F	19	84	20	19	88	19	19	86	20	19.00	86.00	19.67
7	F	15	72	12	13	69	14	13	66	9	13.67	69.00	11.67
31	M	16	79	19	14	73	16	14	69	16	14.67	73.67	17.00
37	M	16	46	17	16	44	13	12	42	17	14.67	44.00	15.67
97	F	20	97	17	20	95	17	20	94	18	20.00	95.33	17.33
9	M	11	56	11	6	49	11	8	43	12	8.33	49.33	11.33
33	M	4	29	2	3	28	4	4	14	3	3.67	23.67	3.00
51	M	20	82	16	19	71	14	20	66	14	19.67	73.00	14.67
10	M	10	60	8	10	59	9	10	56	9	10.00	58.33	8.67
104	M	15	68	12	11	60	10	12	67	10	12.67	65.00	10.67
53	M	17	63	13	14	64	14	13	64	14	14.67	63.67	13.67

Control Class 1 - Posttest (n = 27)

Student Subject Code	Sex	Scorers' Ratings: Scorer 1 SAM Part I	Scorer 1 II	Scorer 1 III	Scorer 2 SAM Part I	Scorer 2 II	Scorer 2 III	Scorer 3 SAM Part I	Scorer 3 II	Scorer 3 III	Performance Scores: SAM Part I	SAM Part II	SAM Part III
64	F	15	69	11	12	63	12	15	64	11	14.00	65.33	11.33
35	F	20	86	15	19	73	12	19	73	12	19.33	17.33	13.00
102	M	19	65	11	18	67	11	18	66	10	18.33	66.00	10.67
25	M	20	94	19	20	91	17	20	91	18	20.00	92.00	18.00
91	F	20	92	20	20	79	15	20	85	17	20.00	85.33	17.33
98	M	17	58	11	11	45	12	16	55	12	14.67	52.67	11.67
12	M	20	76	16	19	76	10	20	72	10	19.67	74.67	12.00
82	M	19	68	12	16	62	11	17	67	11	17.33	65.33	11.33
56	M	8	34	15	6	39	9	8	40	9	7.33	37.67	11.00
43	M	5	45	18	4	44	15	7	44	13	5.33	44.33	15.33
27	F	13	58	11	11	53	10	10	58	11	11.33	56.33	10.67
17	F	18	87	14	19	83	12	19	77	12	18.67	82.33	12.67
79	M	7	43	16	4	36	10	7	36	9	6.00	38.33	11.67
81	M	18	95	17	18	85	16	18	90	17	18.00	90.00	16.67
93	F	10	49	18	8	49	17	8	64	17	8.67	54.00	17.33
58	F	5	46	11	4	42	11	4	37	10	4.33	41.67	10.67
20	F	11	71	16	9	72	14	12	70	14	10.67	71.00	14.67
77	M	12	79	13	7	61	13	14	73	15	11.00	71.00	13.67
84	F	1	65	15	0	57	10	0	66	11	.33	62.67	12.00
26	M	20	92	20	19	77	17	20	88	19	19.67	85.67	18.67
62	M	20	96	20	18	96	18	19	89	19	19.00	93.67	19.00
14	M	2	48	12	1	43	10	1	42	9	1.33	44.33	10.33
41	M	17	78	17	12	64	15	14	69	16	14.33	70.33	16.00
70	M	16	62	16	13	60	15	16	64	14	15.00	62.00	15.00
36	F	19	63	10	15	53	10	17	61	8	17.00	59.00	9.33
16	M	16	57	14	13	60	11	13	61	11	14.00	59.33	12.00
47	M	20	78	20	20	64	18	20	72	19	20.00	71.33	19.00

Control Class 2 - Posttest (n = 25)

Student Subject Code	Sex	Scorers' Ratings: Scorer 1 SAM Part I	II	III	Scorer 2 SAM Part I	II	III	Scorer 3 SAM Part I	II	III	Performance Scores: SAM Part I	SAM Part II	SAM Part III
24	M	13	82	15	13	81	12	13	75	10	13.00	79.33	12.33
19	M	20	96	20	20	86	16	20	80	17	20.00	87.33	17.67
38	M	2	49	12	2	43	10	2	44	11	2.00	45.33	11.00
15	M	3	28	14	2	34	9	2	28	9	2.33	30.00	10.67
55	M	20	97	19	19	86	15	20	85	19	19.67	89.33	17.67
49	M	20	68	17	19	60	15	20	72	17	19.67	66.67	16.33
5	M	20	79	15	20	71	14	20	78	17	20.00	76.00	15.33
21	F	6	40	11	5	37	13	6	40	12	5.67	39.00	12.00
57	M	0	39	11	0	39	10	0	31	4	0	36.33	8.33
29	M	15	63	12	9	57	8	14	64	10	12.67	61.33	10.00
67	F	18	94	20	20	93	18	19	90	19	19.00	92.33	19.00
96	F	20	96	19	20	88	15	20	95	17	20.00	93.00	17.00
99	M	16	58	17	18	60	16	19	58	13	17.67	58.67	15.33
40	M	11	38	15	8	33	9	9	19	9	9.33	30.00	11.00
32	F	20	98	20	20	99	20	20	97	20	20.00	98.00	20.00
34	M	4	39	8	4	41	8	2	31	6	3.33	37.00	7.33
65	F	15	60	15	16	54	10	17	57	10	16.00	57.00	11.67
95	M	4	40	12	3	35	11	2	18	6	3.00	31.00	9.67
105	M	14	75	11	11	65	9	14	65	12	13.00	68.33	10.67
103	M	6	73	16	5	62	10	5	64	13	5.33	66.33	13.00
63	M	13	54	11	12	46	10	14	53	10	13.00	51.00	10.33
54	M	2	39	11	0	36	9	2	30	8	1.33	35.00	9.33
28	M	13	43	12	8	43	10	12	49	11	11.00	45.00	11.00
92	F	3	26	12	3	28	10	2	29	0	2.67	27.67	7.33
89	M	4	33	11	1	37	9	1	37	7	2.00	36.33	9.00

Appendix I

Raw scores for *Primary Measures of Music Audiation*

Experimental Class 1 (n = 21)			Experimental Class 2 (n = 23)		
Student Subject Code	Pretest	Posttest	Student Subject Code	Pretest	Posttest
74	32	34	23	37	34
86	21	34	90	32	34
78	30	37	73	33	37
101	29	31	85	30	36
6	34	38	87	30	34
100	35	35	4	30	39
66	37	37	2	32	39
45	34	35	42	32	36
50	29	34	80	35	36
71	30	35	61	30	37
39	33	35	59	37	37
8	38	39	3	33	27
48	32	34	44	38	40
76	31	36	94	35	37
46	31	33	7	31	33
69	21	31	31	31	34
60	30	36	97	36	40
52	26	35	9	32	34
18	35	38	33	24	23
11	32	37	51	35	38
1	33	37	10	33	33
			104	33	37
			53	29	35

Control Class 1 (n = 23)			Control Class 2 (n = 22)		
Student Subject Code	Pretest	Posttest	Student Subject Code	Pretest	Posttest
64	29	33	24	31	36
102	35	35	19	36	33
25	38	39	38	33	34
91	34	33	15	23	28
98	35	34	49	29	28
12	31	39	5	31	37
82	32	36	21	23	27
56	33	35	57	30	36
43	31	34	29	34	35
27	35	33	67	32	37
17	26	35	96	28	31
79	35	37	99	35	34
81	38	40	32	37	38
93	31	33	34	32	32
58	30	30	65	28	35
20	34	33	95	30	27
84	27	24	105	32	34
26	27	32	103	26	34
62	37	38	54	30	27
41	26	22	28	33	34
70	27	36	92	29	27
16	30	30	89	30	31
47	26	32			

Appendix J

Correlation Coefficients for *Singing Achievement Measures* Part I, II, III,

Primary Measures of Music Audiation, and *Home Musical Environment Questionnaire*

	PRETEST SAM Part I	Part II	Part III	POSTTEST SAM Part I	Part II	Part III	PMMA Pretest	PMMA Posttest
SAM Part II pretest	r = 0.74							
SAM Part III pretest	r = 0.58	r = 0.75						
SAM Part I posttest	r = 0.77	r = 0.71	r = 0.58					
SAM Part II posttest	r = 0.65	r = 0.87	r = 0.72	r = 0.79				
SAM Part III posttest	r = 0.60	r = 0.71	r = 0.76	r = 0.67	r = 0.81			
PMMA pretest	r = 0.45	r = 0.48	r = 0.26	r = 0.44	r = 0.46	r = 0.36		
posttest	r = 0.37	r = 0.45	r = 0.32	r = 0.45	r = 0.52	r = 0.41	r = 0.57	
HMEQ Total	r = 0.16	r = 0.17	r = 0.18	r = 0.17	r = 0.15	r = 0.21	r = 0.13	r = 0.14
HMEQ Siblings	r = 0.06	r = 0.11	r = 0.12	r = 0.07	r = 0.15	r = 0.19	r = 0.18	r = 0.09