THE EFFECTS OF COMPUTER MUSIC LEARNING ACTIVITIES ON THE TONAL APTITUDES OF CANADIAN STUDENTS

by

ALLAN F. ANDERSON

B.Mus., The University of British Columbia, 1972

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Music Education)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1993

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature) _____

Department of Med.

The University of British Columbia
Vancouver, Canada

Date April 26/93

ABSTRACT

With the intent of learning more about the process of assessing music ability, the purpose of this study is to investigate the effects of music learning on music aptitude scores. The problem of this study is to determine if there is a difference between pretest-posttest tonal aptitude scores, as measured by AMMA, for students who possess high and low levels of tonal audiation ability and who either received specialized audiation training on computer or no specialized audiation training.

Forty-eight Grade 11 and 12 music students were administered AMMA as a pretest. An intact music class of 24 students received 13 weeks of computer instruction. The experimental treatment consisted of a computer assisted software program, *Tonal Syntax Tutorial*, which provided audiation practice for high school and college students. A randomly selected group of 24 students received their normal classroom music instruction. Pretest AMMA scores were used as the criterion measure. At the end of 13 weeks, all students were re-administered AMMA as a posttest.

AMMA pretest and posttest *Tonal* scores were organized into a multidimensional design. A covariate analysis of the AMMA scores was calculated and a MANOVA was employed to determine differences between the pretest and posttest AMMA *Tonal* scores. Main effects, interaction effects, and simple main effects were tested at the .05 level of significance. The researcher found no significant difference between the treatment and control group tonal aptitude scores, however, there was a significant difference between levels of aptitude. It was interpreted that the difference between the students who possessed high tonal aptitude and students who possess low tonal aptitude

was not a real difference because the difference in student tonal aptitude levels actually existed before the study began.

The researcher believes that AMMA can be a useful instrument in the ssessment of music abilities of high school students. Also, based on our present knowledge of computer assisted music instruction, it seems that that type of instruction alone is not sufficient to affect a change in tonal audiation ability of high school students.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

APPENDIXES

LIST OF TABLES

LIST OF FIGURES

Figure Page

ACKNOWLEDGEMENTS

The writer wishes to thank his wife, Lorill, and his boys, Cole and Shane, who relinquished valuable family time to allow me the freedom of academic discovery and enlightenment. Enlightenment came through my advisor and mentor Dr. Peter Gouzouasis who revealed a world of music psychology that I had been unaware of in my years of teaching in the public school system. Thank you to my other committee members, Dr. Charles Curtis and Sandra Davies, for their efforts to keep this research relevant and practical for other music educators to study and critique in the future. Dr. Allen Clingman deserves my gratitude for his continuous support throughout my career, as well as Dr. Dennis Vincent who was invaluable in his statistical support and expertise. I would sincerely like to thank Dr. Steven Estrella for allowing me the opportunity to share and expand on his research through the use of his computer learning program. Finally, I would like to thank the music students of Matthew McNair Senior Secondary who so willingly agreed to participate in this study.

CHAPTER ONE

INTRODUCTION, PURPOSE, AND PROBLEM

One of the purposes of psychological testing is to assess cognitive abilities of human beings. Many researchers believe that psychological tests are reliable measures of innate capacities and hereditary predispositions. One may evaluate innate music abilities through measures of music aptitude. Music aptitude may be defined as music potential. Many music educators make judgements on student music abilities and talents based on subjective evaluations and unreliable tests of music aptitude and achievement. Moreover, most music educators rarely evaluate music aptitude and infrequently evaluate music achievement with well constructed objective tools. Although aspects of heredity enter into all human behaviour, it may be unrealistic to expect the psychological development measurement of inherited abilities to be independent of environmental influences. Environmental influences play a role in the development of individual achievement and skills, and these differences are influenced by early childhood experiences.

A major criterion for the consideration of a test as a measurement of music aptitude is the ability to generalize about fundamental abilities and basic skills. Ideally, to take an aptitude test a student need not possess specific skills. That is because, skills are learned over time and are thus related to achievement. One threat to the construct validity of IQ tests is that in order to write an IQ or academic intelligence test, it is assumed that the child to some extent can read and write in a language. Linguistic skills are learned; therefore, to some extent those tests are a measure of linguistic achievement. Some contemporary music

aptitude tests (Gordon, 1965, 1979, 1982, 1989) differ from IQ tests in that children do not need prior music knowledge in order to write them.

Throughout history, music psychologists have attempted to discover whether music aptitude is either a product of nature, which can be described in terms of biological and hereditary factors, or of nurture, which can be described in terms of environmental influences and early learning experiences. It was not until the late 18th century that German psychologists became interested in the nature of music aptitude. European researchers such as Pear (1911) and Revesz (1953), as well as the American researcher Stanton (1922), believed that music aptitude was innate, a product of nature. Seashore (1919) believed that music aptitude was both innate and inherited, both biological and genetic. The philosophical premise that music aptitude is either innate or inherited has never been confirmed through research, but the notion of music aptitude as an innate construct is consistent with research on intelligence in general psychology (Gesell, 1929; Piaget, 1952). While music aptitude is not specifically discussed by developmental psychologists, they do elude to similar concepts in their cognitive-stage models and through their perspectives of change in mental structures. As of yet, research designs created to assess the role of heredity in the aptitude of children have not been successful in the evaluation of this factor (Stanton, 1922). Edwin Gordon, in his book *The Nature, Description, Measurement, and Evaluation of Music Aptitudes* (1986, p. 7-8) writes that music aptitude is a product of both nature and nurture. That is, we are all born with a degree of music aptitude, but unless this aptitude is nurtured at an early age, it will decrease. However, based on existing music research literature, he believes that music ability is innate but not necessarily inherited.

The first published standardized test of music aptitude was the *Seashore Measures of Musical Talent* (Seashore, 1919; revised, 1939). As late as 1957

Seashore continued to update the test. The test is designed for students, age 10 through adult. In designing this test, Seashore used a reductionist or atomistic approach to music aptitude; he believed that music is comprised of different elements that work as one, but that can be measured separately.

From a Gestalt perspective, the major problem with Seashore's test is that it only measures the individual components--pitch, loudness, time, and timbre--of music on an atomistic level. Pitch, loudness, time, and timbre are used to describe the actual physical aspects of sound. Seashore believed that pitch aptitude is the ability to discriminate between two separate tones rather than the ability to discriminate between two tonal patterns or two melodic motives. These two tones are performed at approximately 500 cycles per second and vary by no more than 10 cents. The cent is based on the equally tempered scale with 1200 cents equal to one octave. An interval of a semitone contains 100 cents, therefore, the pitch variance which Seashore uses (10 cents) is only one tenth of a semitone.

For Seashore, loudness aptitude is defined as the ability to discriminate whether the second tone, in terms of volume, is either softer or louder than the first tone. The variance in loudness, as measured in decibels, has a range from 0.5 to 4.0 decibels. Seashore uses the decibel as a comparison between, or ratio of, two quantities of loudness. For Seashore, time aptitude is defined as the measurement of length, longer or shorter, between the pairs of tones and is measured in seconds. The variance of time between the performances of tones ranges from 0.5 to 0.3 of a second. For Seashore, timbre aptitude is a measure of a student's ability to decide whether two pairs of notes are the same or different as differentiated by possible changes to the third and fourth harmonics. The overtones or harmonics of each instrument are different and unique in the world of sound. Timbre is that quality of sound which allows one to distinguish

between two sources, e.g., a clarinet or an oboe, which produce sustained sounds of the same pitch.

Because of the way that Seashore defined and measured aptitude in terms of pitch, loudness, time, and timbre rather than in terms of tonal, rhythmic, melodic, and harmonic elements of music, Seashore was labelled an atomistic researcher by those who elaborated in a Gestalt approach to music aptitude. Seashore's subtests of pitch, loudness, time and timbre focus only on the acoustic components of music, and ignore the musical elements of tone quality, sense of consonance, sense of dynamics, and the sense of rhythm (Gordon, 1987, p.19-20). Gestaltists believe in the concept of holism, the notion that human beings listen to music as a complete entity, therefore, the measurement of component parts of a melody would not be comparable to the measurement of the whole melody.

The followers of the Gestalt movement in psychological research, led by Mursell (1931) in the United States, believed that music aptitude was linked to intelligence, and that it should not be separated and tested as an individual component of intelligence. In other words, general intelligence is an essential part of music aptitude. They also felt that music aptitude test items should be constructed with music related examples performed by a music instrument. The greatest criticism of Seashore's test is that it actually is a measure of music achievement, because prior music learning is needed to answer many of the questions on the test. For example, the use of the concepts of "high" and "low" to describe pitch change is a learned adult construct. Also, it is primarily related to music notation symbols which are learned by children in the classroom. Whereas reading music involves achievement, most children are able to either sing, perform, or read a simple song without knowledge of the concepts of "high" and "low" and music symbols.

Although their test battery is similar in construction to Seashore's test, Kwalwasser and Dykema (1930) claimed to measure different music constructs with the *Kwalwasser-Dykema Music Tests* (K-D Music Tests). Test items are based on actual melodies that consist of intervals no less than a semitone apart. Rather than using random computer generated melodies, they used music examples taken from traditional European melodies and harmonies. Because the music examples are recognizable, the construct and content validity of the test may be challenged. Regardless, their subtests are considered to have more logical music content than Seashore's test (Schuter-Dyson and Gabriel, 1981).

Kwalwasser and Dykema believed that the addition of four subtests of tonal movement, melodic taste, pitch imagery, and rhythmic imagery would make the test results more valid as a predictor of music aptitude. One of those constructs, melodic taste, is a learned concept that is based on environmental and cultural experiences. Sound may be a universal phenomenon, but the syntax of musical taste is not universal because musical taste is in part determined by exposure to particular music styles. As with the Seashore test, the K-D Music Tests (1930) were designed for students age 10 through adult. Seashore's components of pitch, intensity, time, rhythm, timbre, and tonal memory are also included in this test battery. Similar to Seashore's test, it takes one hour to administer. No reliability data is mentioned in the test manual, but an overall average composite reliability (.40) as reported by Farnsworth (1969) is quite low. Despite the low reliability levels found by Holmes (1954), Whybrew (1962), Lundin (1967), and Farnsworth (1969), the battery was used extensively in the United States from 1930 to the late 1960s. Holmes (1954) revised the tests with a new set of scoring keys and norms which resulted in high reliability coefficients (r=.91) for high school students.

Herbert D. Wing's *Tests of Musical Ability and Appreciation* (1948) consists of seven subtests, three subtests which measure aural acuity, and an additional four subtests which measure music preference. Wing defines aural acuity as the ability to decide whether two performances are different because of a change in either rhythmic accents, harmonization, dynamics, or phrasing. Ages eight to adult may be assessed with this battery of tests which takes one hour to administer. Wing was a follower of Gestalt psychology in that he believed aptitude tests should only include relevant music examples played by recognizable musical instruments. He defined relevant music examples as chord progressions familiar to the citizens of England and common to English music and English composers. He also thought that general intelligence was an essential element of music aptitude. Unlike Seashore, Wing believed that musical memory included both tonal and rhythm elements in a melodic context. Wing's insightful additions to the notion of music aptitude test construction include the measurement of music preference, the definition of the interaction between tonal and rhythmic components, and the use of tonal music examples and actual music instruments in his aptitude tests. Composite test reliability coefficients are quite high, (r =.86, Bentley, 1955; r =.90, Wing, 1962). Despite these high reliability coefficients, reliability coefficients for children under nine years of age has never been investigated or published. Objective validity is also questionable since longitudinal studies which use Wing's test battery were never conducted. Also, his reliability calculations are based on multi-age groups, hence, the reliabilities are inflated.

Raleigh M. Drake published his *Drake Musical Aptitude Tests* (1954) based on many of the same notions of music memory that were previously advanced by Wing. With his tests, the music aptitudes of people aged eight through adult may be evaluated in 80 minutes. The memory subtest consists of 54 items.

These items are developed from 12 melodies which are used from two to 12 times. The second example in the pair is either the same or different because of a pitch, rhythm, or modulation change. The rhythm subtest consists of 50 atomistic test items. These test items are considered atomistic because rhythm aptitude is measured without reference to pitch. Published reliability coefficients range from (r =.85 to .93) for the memory subtest and (r =.56 to .96) for the rhythm subtest (Tanner and Loess, 1967). With this wide a range of reliability coefficients, one must conclude that the low validity is a result of a test which is not stable on a reliability level. One of the problems with this test is that Drake used music examples that may be recognized by some students, which means that prior learning might affect test results. As with the Seashore and Wing aptitude tests, research has shown that Drake's test measures achievement, not aptitude (Gordon, 1958).

The *Bentley Measures of Musical Abilities* (1966) is designed for students aged seven through fourteen and takes 20 minutes to administer. The battery includes four subtests: Pitch Discrimination, Tonal Memory, Chord Analysis and Rhythm Memory. Like the music psychologists who preceded him, Bentley used microtonal patterns (i.e., the discrimination of minute pitch fluctuations in individual pitches) in his pitch discrimination subtest. Also, tonal memory and rhythm memory are tested separately. The separate evaluation of those two concepts in two subtests made the test easier for young students (Schuter-Dyson and Gabriel, 1981, p.29). Reliability coefficients for pitch, tonal memory, chord analysis, and rhythm memory were (r =.74, r =.53, r =.71, and r =.57) respectively (Bentley, 1966). Composite reliability coefficients for this test range from r =.84, (Bentley, 1966) to r =.60, (Rowntree, 1970). Overall reliability of this test is questionable since the only published total test scores and analyses are for a heterogeneous group of Grade 5 through Grade 7 children.

The reliability coefficients are inflated because of the wide range of student ages, hence music experience, thus increasing the variability in the test scores. Test balance is also a concern because the battery includes 20 chord analysis subtest items but only 10 memory subtest items. No longitudinal and predictive validity studies that use the *Bentley Measures of Musical Abilities* exist in the research literature.

In 1965, Edwin Gordon published the *Musical Aptitude Profile* (MAP). This battery of seven subtests is considered by many music educators to be the best diagnostic standardized music aptitude test available today. The battery may be used to evaluate Grade 5 through Grade 12 students. In the test manual, Gordon provides separate item analyses and reliabilities for each grade level. The MAP is designed to be administered over three days, with the objective, non-preference subtests of *Tonal Imagery (Melody-T1* and *Harmony-T2)* and *Rhythm Imagery (Tempo-R1* and *Meter-R2)* each composed of 40 questions, and the three preference subtests of *Musical Imagery (Phrasing-S1*, *Balance-S2* and *Style-S3)* each composed of 30 questions. A violin and a cello are used to perform all the music examples. An "in doubt" response is included on the answer sheet and students are directed to use this response instead of guessing answers. Gordon feels that this response option adds to the high validity of MAP (Gordon, 1986).

Over 180 studies (Scheuter, 1991) have used the MAP as a criterion measure. All of those studies verify the high reliability and validity of the MAP. In the manual, Gordon states that MAP is a test of stabilized aptitude. Based on research (Fosha, 1960; Gordon, 1968; Tarrell, 1965), he theorizes that music aptitude stabilizes at age nine. He also believes that music aptitude is normally distributed throughout the population. Contemporary research, while subjectively questioned by some, seems to verify those conclusions (DiBlassio,

1984; Flohr, 1981; Gordon, 1979; Gouzouasis, 1987; Jessup, 1984; Levinowitz, 1985).

Early in his career, Gordon believed that stabilized music aptitude should be described in both Gestalt and atomistic principles (Gordon, 1971, p.22). All of Gordon's more recent music aptitude tests, *Primary Measures of Music Audiation* (PMMA, 1979), *Intermediate Measures of Music Audiation* (IMMA, 1982) *and Advanced Measures of Music Audiation* (AMMA, 1989) have one major theme, which is the overriding importance of the concept of audiation. Audiation is the ability to conceptualize music when sound is not physically present. Thinking is to language what audiation is to music, a complex conceptual process (Gouzouasis, 1992). Gordon's (1989) model includes seven applications of audiation--the abilities to listen, to read, to write, to perform, to recall and notate, to create, to improvise and perform, and to compose music either in one's mind or on paper.

Gordon (1989) has attempted to theorize the steps that occur when we audiate familiar and unfamiliar music. Stage one involves the retention of a series of notes for only a few seconds. Stage two involves the conscious organization of a series of notes into patterns of essential notes. In Stage three one consciously retains in audiation the pattern of essential notes that were just organized. Stage four involves the comparison of known patterns of essential notes with the ones that were just audiated. In Stage five, one predicts patterns through association, of essential notes before they are heard. Stage six involves the prediction of patterns that are about to be heard.

In 1979, Gordon published the *Primary Measures of Music Audiation* (PMMA). This test of developmental music aptitude is designed for children in Kindergarten through Grade 3. To take a standardized music aptitude test such as MAP (Gordon, 1965) and PMMA (Gordon, 1979) a child needs no prior

instruction in music, but does need previous listening experiences. To write either MAP or PMMA, a student only need understand the meaning of the concepts "same" and "different." Based on longitudinal research (Gordon, 1986; Woodruff,1983), it has been determined PMMA is the only valid music aptitude test available for children ages five though nine.

The short administration time for each of the two subtests, 18 minutes each, is a very positive aspect of this test. Most important, children do not need to know how to read or write language. Questions on the answer sheets are identified by pictures, not numbers or words. Children are asked to listen to and respond to two musical examples, and to decide whether they sound either the same or different. If they are the same, the child circles the two happy faces; if they are different, the child circles the one happy and one sad face. Five seconds are provided between each test item to allow students time to audiate and mark their score sheets.

Gordon believes that unless children receive positive early musical experiences, they will not be able to maintain the level of aptitude with which they were born (Gordon, 1979). In other words, children cannot surpass the level of music audiation with which they were born, but they can decrease in their ability to audiate if they are not provided with quality music experiences. Because of the interaction of audiation abilities and environmental variables, either more or less music instruction and exposure to music, children birth through age nine are considered to be in the developmental stage of music aptitude. Gordon states:

> It cannot be over-emphasized that early informal and formal instruction, particularly from birth to age nine, is of greater consequence than formal instruction in music after age nine. The more appropriate early experience and formal instruction in music are, the higher the level at which a student's music aptitude will

> stabilize. Also, the younger a child is when he begins to receive informal and formal instruction in music, the more he can profit from such instruction and the higher level at which his music aptitude will eventually stabilize. (Gordon, 1987, p.9).

In 1982, Gordon published the *Intermediate Measures of Music Audiation* (IMMA), which is designed to more precisely examine the music aptitude of kindergarten through grade four students The only difference between PMMA and IMMA is that IMMA is constructed with tonal and rhythm patterns that are more difficult to audiate and discriminate, hence, one is able to more accurately evaluate children who score higher than the 80th percentile on PMMA. Like PMMA , IMMA is given over two successive days with tonal and rhythm elements evaluated separately. As in all of Gordon's manuals, interpretation of results from pilot studies and the standardization study, research implications, and suggestions for future formal and informal instruction are well documented.

Gordon's *Advanced Measures of Music Audiation* (AMMA, 1989) was originally designed for undergraduate and graduate students in college music programs. Despite this initial categorization, it has been successfully used to evaluate the audiation abilities of children as young as Grade 5 (HEADS, 1986-87). Standardized scores and percentile rankings are available in the test manual for high school students, college music majors and college non-music majors. *Tonal*, *Rhythm*, and *Composite* scores are gathered in only 30 minutes from this test of 30 questions. Four seconds are provided between test items to allow students time to audiate and mark their answer sheets. Similar to PMMA and IMMA, the test involves the choice as to whether two successive musical examples played on a Yamaha DX-7 keyboard either sound the same, sound different because of a tonal change, or sound different because of a rhythm change. However, whereas PMMA is based on the audiation and discrimination of similarities and differences of tonal and rhythm patterns in

separate contexts, AMMA is based on the audiation of entire melodies and the discrimination of similarities and differences of tonal and rhythmic aspects of entire melodies. Each test item was derived from a computer software program. Because AMMA is relatively new, no longitudinal studies that use AMMA have been conducted.

Estrella (1992), in his study of music aptitude, measured the effects of tonal training and practice on AMMA scores. A computer assisted software program designed for individualized instruction was used for the study. Estrella composed music patterns similar to those found in AMMA, and trained students with those patterns in an attempt to improve AMMA posttest scores. A computer was used because a subroutine of the instructional program was able to diagnose and control the speed in which a student progressed, based on their demonstrated skills. The use of a computer also eliminated the possibility of teacher bias. In his study, which included a comparison of two experimental groups who learned tonal pattern and tonal function skills and two control groups, Estrella found that AMMA scores could be improved ($p<.05$) with specialized training.

One may interpret that the increase in audiation ability was not necessarily a measure of an actual increase in developmental music aptitude but merely a measure of a child's ability to activate and retrieve dormant audiation abilities (Gouzouasis, 1992). Estrella's findings are contrary to those of Gordon (1990), which poses some interesting questions with regard to the diagnostic validity of AMMA. It is with those perspectives in mind that the researcher chose to explore the issue of AMMA as a measure of stabilized music aptitude.

With the intent of learning more about the process of assessing music ability, the purpose of this study is to investigate the effects of music learning on music aptitude scores and tonal skills. The problem of this study is to determine if

there is a difference between pretest-posttest tonal aptitude scores, as measured by AMMA, for students who possess high and low levels of tonal audiation ability and who received either specialized audiation training, and students who possess high and low levels of tonal audiation ability and received no audiation training.

CHAPTER TWO

REVIEW OF THE LITERATURE

There are few published studies which investigate the topic of developmental and stabilized music aptitudes. An *ERIC* (1966) search revealed little information on this research topic. Few studies with pertinent information were discovered through a search of the *Dissertation Abstracts International* and *Comprehensive Dissertation Index* (1966-91). Articles found in *The Quarterly* (Hohn,1991; Colwell and Abrahams, 1991; Schuler, 1991; Walters, 1991; Grunow, 1991; Azzara, 1991; McDonald, 1991; Saunders, 1991; Trusheim, 1991) helped to focus this study, since two successive volumes were recently dedicated to the work of Edwin Gordon. The Estrella (1992) dissertation was easily acquired because the researcher was an active participant in the administration of a treatment effect for that study.

Secondary sources were gathered from selected readings on the topic of music aptitude and how it is affected by training and practice. These sources included journals such as the *Journal of Educational Research* (1965-1991), *The Bulletin of the Council for Research in Music Education* (1967-1991), and the *Psychology of Music* (1972-1991). The book *Psychology of Music Ability* (Shuter-Dyson and Gabriel, 1981) offered valuable historical facts and a discussion of psychological constructs in the study of music aptitude. Also, the many books and articles by and about Edwin Gordon (Gordon, 1965, 1970, 1971, 1979, 1982, 1986, 1989, 1990) were a tremendous resource to learn more about music aptitude in general and in specifics with regard to how it may be measured.

In the present study, research by Flohr (1981), Jessup (1984), Levinowitz

(1987), and Gouzouasis (1987) all used *Primary Measures of Music Audiation* (PMMA; Gordon, 1979) as their criterion measure of music aptitude. The Estrella study (1992) is directly related to the present study in that the same computer instruction program which was used by Estrella is used in the present study. Estrella is the first researcher to study the effects of training and practice on audiation as measured with the *Advanced Measures of Music Audiation* (AMMA; Gordon, 1989). AMMA is primarily designed as a test of stabilized music aptitude for college-age students. Although there are no reliability item difficulty and item discrimination coefficients provided in the test manual, this test has been standardized as a criterion measure with grade 7 through Grade 12 students.

The Flohr Study

Flohr's study (1981) was designed to determine the effect of short-term general music instruction on the developmental music aptitudes of five year old children. Twenty-nine children from the Child Development Center of Texas Women's University participated in the study. Each child was randomly assigned to one of three treatment groups. Prior to instruction, the *Tonal* and *Rhythm* subtests of PMMA (Gordon, 1979) were administered to all of the children.

Children who were assigned to the first experimental group received 12 weeks of general music instruction that emphasized improvisation. Children who were assigned to the second experimental group received 12 weeks of regular general music instruction with no emphasis on improvisation. All children in both groups received their respective instruction from two student teachers twice a week for 25 minutes. The third group was a control group in which children received no music instruction. At the end of instruction, both subtests of the PMMA were administered to all children. The dependent

variable was the PMMA scores .

An analysis of covariance (ANCOVA) was used because the numbers of students who took the PMMA differed for each of the three groups. Pretest scores were used as a covariate because the three groups were sufficiently different on the PMMA pretest scores. There was a significant difference in observed posttest scores for the two groups which received music instruction. T-tests were performed to determine if there was a difference between the pretest and posttest scores for the groups who received music instruction and the control group. Flohr found that children who received music instruction audiated at a significantly higher level than children who did not receive music instruction. In other words, children who received music instruction had a significant increase in audiation ability.

Flohr believes that a child's developmental music aptitude, as measured by PMMA, can be influenced by general music instruction. Flohr also believes that the effects of general music instruction on young children may be temporary. Based on those positions, he felt that further research was needed to study the effects of instruction over time.

In this study, Flohr used a quasi-experimental design with a sample of only 29, five-year-old children. The small size of each group raises the possibility of a Type I error. The design used in this study was appropriate and the methodology fairly clear. To replicate this study one might use a two-dimensional model with more subjects. The two-dimensional design would be comprised of levels of instruction and high and low levels of aptitude. A two factor analysis of variance (treatments by levels of aptitude) might be used to analyze the effects of general music instruction for children who possess high and low music aptitude The practical outcomes of this replication would be to emphasize the importance of music instruction for all young children in the

public school system.

The construct of time is an important variable that should be included in future replication studies that examine developmental music aptitude. Because development is measured over time, questions regarding length of instruction (in weeks) and amount of instruction per week (days and minutes) must be considered in future research (Gouzouasis, 1992).

Levels of developmental music aptitude should also be considered as an independent variable in future studies. To analyze the effect of instruction on children who possess various levels of music aptitude would help music educators to be more diagnostic in their approach to music instruction. Children who possess low rhythm aptitude or low tonal aptitude may be given instruction appropriate to their abilities, without loss of self-esteem, while students who possess high rhythm aptitude or high tonal aptitude may be given instruction to challenge their audiation abilities and keep their interest.

The Levinowitz Study

Levinowitz (1987) studied the effects of song instruction, which included songs with and without words, on the developmental music aptitude and singing achievement of children in kindergarten and first grade. Before instruction began, three kindergarten classes and three-first grade classes were administered PMMA (Gordon, 1979). All children received one of three types of song instruction for one academic year. One group learned to sing songs primarily with words (T1). They heard no more than two songs without words sung during any class period. The second group learned to sing songs primarily without words (T2). For them, no more than two songs were sung with words during any class period. The third group learned to sing all songs with words (T3). At the end of a year of instruction, PMMA was administered to all of the children. Also, to evaluate singing achievement, all children were

individually tape-recorded singing two criterion songs. The two criterion songs were taught in the last month of instruction. One song was in major tonality and triple meter and the other song was in minor tonality and duple meter.

Two separate three-dimensional designs were used to determine the effects of singing instruction with and without words on the PMMA *Tonal* scores and the PMMA *Rhythm* scores of the kindergarten and first grade children. For the kindergarten children, two three-way analyses of variance (treatments by levels of high and low PMMA aptitude scores by one year of instruction time) with repeated measures on the time factor were performed on the PMMA *Tonal* and *Rhythm* subtest scores. The first grade children's scores were analyzed separately using this same technique. Singing achievement scores were organized into two, two-dimensional designs - - (treatments by levels of music aptitude), one for tonal achievement scores and one for rhythm achievement scores. As with the aptitude scores, the kindergarten children's and first grade children's achievement scores were analyzed separately. Two, two factor analyses of variance were separately performed on the rhythm and tonal achievement scores for each of the kindergarten and first grade children. This design was used to determine the comparative effects of treatments on the singing achievement of kindergarten and grade one children (levels).

Statistically significant two-way interactions between high and low levels of aptitude and amount of instruction time were found for the tonal analysis of kindergarten children and for both the rhythm and tonal analyses of first grade children. Also, a significant three-way interaction was found in the rhythm analysis of kindergarten children.

Levinowitz found that the audiation ability of young children with low developmental music aptitude is enhanced by song instruction, regardless of whether words were used. She also believes that tonal and rhythm music

aptitudes scores are not affected by song instruction. Singing achievement scores were also not affected by the three different song instruction methods. The one ambivalent discovery for Levinowitz was that the song instruction which was comprised of only words increased the audiation skill of children with high rhythm developmental music aptitude. She concluded that the rhythm of the words in songs may have helped these students to audiate and sing the rhythm of the music more precisely.

In this study, all of the groups were randomly selected. The description of the sample is adequate for anyone who wishes to undertake a replication study, although a further description of the ethnic and socio-economic background of the children might be helpful to future researchers. The design, methodology, and analyses are appropriate for this topic. The use of high and low levels of music aptitude as an independent variable, in contrast to Flohr's study, is a very practical aspect of this study. Also, the length of instruction time for this study was longer, one year versus 12 weeks, than the Flohr study.

It is interesting that Levinowitz chose to have the criterion songs kept in their original format, that is, the songs with words were not sung without words and the songs without words were not sung with words. It would be of interest to this researcher to know how the instruction methods might have compared if the children had all sung the same songs during their class periods. Music information such as tonal and rhythmic complexity of the songs that were sung during the year is very unclear.

The conclusion that children who possess high rhythm music aptitude might benefit more from song instruction that uses only words needs further research. Questions to be considered might include to determine if the music helps students to learn the words of a song or if the words help the student to learn the music. It would be interesting to see if there is a high correlation between IQ

levels and music aptitude levels in the grade one group that was investigated. The published research by Gordon stipulates otherwise.

Most important, the findings of this study suggest that developmental music aptitude levels may be increased through a variety of techniques for kindergarten and grade one students who possess low tonal and rhythm aptitudes. The belief that the PMMA scores of children with high aptitude levels are unaffected by any one method of instruction reinforces the validity of using PMMA for the measurement of music aptitude.

The Jessup Study

Jessup (1984) investigated valid ways for music supervisors to evaluate music teachers. The problems of this study were to determine the comparative effects of indirect and direct music teacher behaviour upon overall developmental music aptitude scores, music listening achievement, and the performance of rote songs and tonal and rhythm patterns of students who possess high and low levels of developmental music aptitudes. The terms "indirect and direct teacher behavior" were used to describe the two methods of teacher classroom instructional strategies. Indirect teacher behaviours allow students more freedom to respond than do direct teacher behaviours. In this study, Jessup attempts to prove that standardized music aptitude tests can be used in conjunction with non-standardized achievement tests to find results that will be useful to music supervisors for measuring and evaluating music teacher effectiveness.

Jessup used a system designed by Flanders (1960) to observe behaviour interactions among students and teachers. Flanders observed that Grade 8 geometry and socials studies students scored higher on written achievement tests (Flanders, 1960, p.2) if the teacher used either indirect teaching behaviours or techniques. An indirect teacher behaviour includes the praise of

students, use of student ideas, and the use of music echo responses. In that context, students are constantly stimulated and encouraged to achieve. A direct behaviour teacher will lecture, give directions, criticize, and play music for listening. No effort is made by teacher direct behaviour to either interest students in music or to motivate them to achieve in music.

A sample of 85 students enrolled in four intact heterogeneous (academic ability) Grade 2 classes from the same elementary school in Montgomery County, Pennsylvania participated in the study. Students were from a predominantly middle class background. All students were administered the *Tonal* and *Rhythm* subtests of PMMA (Gordon, 1979) as a pretest, in successive weeks. Each class met for one 30 minute period of instruction per week. The researcher trained the school music specialist on direct and indirect instruction techniques. The researcher and the school music specialist each taught two classes. Students taught by the researcher received 13 weeks of instruction while the students taught by the music specialist only received 11 weeks of instruction. Illness and school holidays were given as reasons for the disparity in instruction time. Each teacher used primary indirect teacher behaviour for treatment group one and primary direct teacher behaviour for treatment group two. Four music teaching objectives were selected from the *Silver Burdett Music Series* , *Book Two (1980).* These objectives included beat (no beat), tempo (fast-slow), dynamics (loud-soft), and register (high-low).

A cassette was made of each lesson to ensure that direct and indirect behaviours were used. Two five-minute tape samples of each of these classes were analyzed by two independent judges. The judges used Jessup's *Music Classroom Instruction Analysis* (*MCIA*, 1978) system. This system of observations was based upon Flander's system (1960) and modified by Jessup for applications in the music classroom. The "M" letters represent Jessup's

additions to Flander's (1960) categories. The 10 Flanders questions and the additional five Jessup questions include 1) accepts feelings, 2) praises and encourages, 3) accepts or uses ideas of students, M3) accepts or uses musical ideas of students, 4) asks questions, M4) asks musical questions, 5) lectures, M5) lectures through musical performance, 6) gives direction, 7) criticizes or justifies authority, 8) student talk--response, M8) student talk--prescribed musical response, 9) student talk-initiative, M9) student talk--creative musical response, 10) silence or confusion.

The four criterion measures used included PMMA (Gordon, 1979) *Tonal* and *Rhythm* posttest scores, the *Silver Burdett Music Competency Tests* (SBCT; Colwell, 1980) that corresponded to the four teaching objectives, student performances of two phrases of a rote song taught to all students in the last three weeks of instruction, and student performances of four unfamiliar major and minor tonality patterns and unfamiliar duple and triple meter patterns. Two, five-point continuous rating scales were constructed for each tonal and rhythm pattern. The patterns were chosen by the researcher.

Interjudge reliability of the rating scale was investigated for the 80 individual 30-minute class meeting samples, 20 for each of the four classes. Overall scores for all 10 meetings of each of the four separate classes were correlated. Interjudge reliability also was investigated for each of the 15 interaction analysis categories, the total direct categories, and the the total indirect categories. The extent of indirect and direct verbal behaviours demonstrated by each teacher (A and B) was investigated by calculating two types of indirect/indirect verbal behaviour ratios for each of the 80 individual 30 minute class meeting samples tallied by the judges. Scores were organized into four, three dimensional designs (treatments by levels by replications). The design was used once for each of the four criterion measures. The 66th percentile was used for high and

low ranking of PMMA (Gordon, 1979) student scores in order to achieve equity of cell size. Treatments were direct/indirect teacher behaviours, levels were high and low levels of music aptitude based on composite PMMA (Gordon, 1979) scores, and replications were the two teachers A and B. Interaction effects, main effects, and simple effects were calculated. For all analyses the level of confidence used for statistical significance was (.05). A three-way analysis of variance was used to analyze the data.

Interjudge reliability for the 80 classes ranged from (r=.77 to .99). Teacher B (the researcher) was more successful in effecting a difference between indirect and indirect teacher behaviours. PMMA (Gordon, 1979) pretest and posttest scores in this study were two to two and a half standard errors higher than those published in the test manual. Only composite scores were analyzed for tonal, rhythm, and rote songs. The mean and standard deviations of student scores on the SBCT were not compared to those in the test manual since only four of the six subtests were administered to the students. The entire test contains 20 items, but only 15 items were used in the music achievement test. Although no statistically significant differences were found among the PMMA gain scores for the indirect/direct treatments, a significant main effect was found for developmental music aptitude. Students with low pretest aptitude scores increased their scores more than did students with high pretest scores. The observed mean for teacher A's direct behaviour class was higher than teacher B's direct behaviour class, while the observed mean for teacher B's indirect behaviour class was higher than teacher A's indirect behaviour class. It was later discovered by Jessup that teacher A's direct behaviour class contained three gifted children and an intact high reading level group of students. Therefore, this was not an academically heterogeneous group. Jessup also observed that there was a high correlation between IQ and music achievement

when the SBCT were used.

Several of Jessup's conclusions need examination in relation to the present study. First, the lack of significant treatment effects may be due to the fact that Jessup's methodology and objectives were not totally understood by either the other music specialist or by the judges who rated the tapes. The lack of similarity found between the two teaching styles shouldn't have affected the indirect/direct results if the behaviours had been totally understood by the other music teacher. If the behaviours were totally understood by the other teacher, then the external validity of this study is suspect. The issue of amount of instructional time is a very important aspect of this study, in that one 30 minute class a week may not be enough to affect student behaviours, attitudes, or skill development. More music instruction time may be needed to produce a higher level of achievement for elementary students. For the present study, a computer assisted instruction program will eliminate the possibility of teacher error or bias. Also, two sessions a week in the present study may have more of an effect on learned skills.

One of Jessup's informal observations was that students with high IQs scored higher on the tonal pattern exercises. That is not surprising, in consideration that the method of data measurement and the SBCT test design itself all relate to the measurement of music achievements, not the measurement of music aptitude. The students who scored low on the PMMA pretest, but demonstrated the most significant improvement on the PMMA posttest, most likely had not been given the opportunity to discover their music potential through previous training and instruction in music. Since the reader is not given data on students' music backgrounds, one might accept the possibility that the students who scored low on the PMMA may not have had any previous training and instruction in music. This study does collaborate contemporary

belief that scores on PMMA (Gordon, 1979) may be improved both normatively and idiographically. A long term study of these students would have been useful to study the effects of skill and practice on standardized aptitude scores.

This writer does not believe the labels of "direct and indirect" teaching are appropriate, because, although the terms seem appropriate for this study, they are too vague and misleading. Observations of "direct and indirect" teaching are quite subjective, but when the data is presented in this objective format, one might question the validity of the SBTC test design.

The Gouzouasis Study

Gouzouasis's study (1987) was designed to learn about the role of hot (electric guitar informal song and movement accompaniment), warm (acoustic guitar informal song and movement accompaniment), and cool (non-accompaniment informal song and movement) media, from a McLuhanistic perspective (McCluhan, 1964), in the acquisition of music skills and in the development of tonal aptitude of preschool children. The effects of the different levels of accompaniment technology were explored. Seventy-eight five-year-old preschool children were assigned randomly to one of three instruction groups. The children who participated in the study were selected from an urban, middle-class population in Philadelphia.

The *Tonal* subtest of PMMA (Gordon, 1979) was administered as a pretest to all children. The investigator taught the three groups of children for 14 weeks in two nearby locations of the same community day care center. Each group of children met twice a week for 30 minute sessions of music instruction.

The *Tonal* subtest of PMMA was readministered to all of the children after the 14 weeks of instruction. Children's singing achievement was rated by two independent judges using the same five-point continuous rating scale for each of the four criterion songs in major, minor, dorian, and mixolydian tonality. To

determine the effects of the treatments on the children's rote singing achievement, the combined judge's ratings were organized into four, two-dimensional designs for the major, minor, dorian, and mixolydian criterion songs. Singing achievement scores of the children were divided into high and low levels of developmental tonal aptitude by using the pretest *Tonal* scores on PMMA as the basis for levels. Combined judge's ratings, organized into four, two-dimensional designs of the criterion songs were used to determine the effects of the accompaniment treatment on singing achievement. *Tonal* pretest and posttest scores were organized into a two-dimensional design to study the treatment effects on developmental aptitudes of the three classes.

No significant interaction or main effects for treatment were noted for the major, minor, or dorian criterion songs. Only the mixolydian criterion song had a significant treatment effect.

Gouzouasis observed that there was a significant gain in developmental tonal aptitude for children who participated in the study--children who possessed low tonal aptitude gained more audiation ability than children who possessed high tonal aptitude. Based on non-significant but observable differences in mean singing achievement scores, he believes that song singing skills seem to be affected by the types of accompaniment media, but not in any consistent pattern. Because of these inconsistencies, he recommends that all types of accompaniment should be used to teach songs.

In this study, Gouzouasis used a quasi-experimental design which incorporated three treatment groups. Since the researcher taught all of the classes, and the researcher is a proficient guitarist, instructional reliability was not a concern. The research design is described in enough detail to allow replication of this study. All the criterion measures are described in detail and are easily understood. Interjudge reliability for the rating scales was high. That

was because the researcher and an elementary music specialist were the judges, and because the researcher designed an evaluation kit (included in the thesis appendix) that clearly explained how to use the rating scale. Gouzouasis reported interjudge reliabilities for the criterion major, minor, mixolydian, dorian, and all songs combined were (r = .94, .92, .87, .96, and .97), respectively (p.33).

The use of the terms "hot, warm, and cool" in describing the treatments might be confusing for some readers, since Gouzouasis explains and extends the common meanings of these terms in a theoretical model. The model is an attempt to relate McCluhan's theory of communication to Gordon's theory of music learning (1990). The statistical analysis used is very powerful for analyzing individual differences in singing achievement using major, minor, dorian, and mixolydian songs. In hindsight, instead of the several ANOVAs that were used to analyze the criterion songs, a multivariate analysis of variance might have been more appropriate for this study, since it would have involved less calculations and thus lessened the chance of a Type I error. One may hypothesize that his choice of analysis was influenced by the notion that each tonality has its own form of syntax and that at the time when the study was conducted an atomistic perspective of tonality was in vogue, hence, each criterion song was analyzed separately.

Gouzouasis concludes with some very important theoretical discussion based on observation about the relationship between developmental music aptitude and song singing skills. He maintains that the Gestalt effect of singing songs in many different tonalities to improve music aptitude must be studied before educators too readily adopt popular elementary teaching techniques which rely heavily on song and tonal patterns in pentatonic tonalities. The influence of mixolydian tonality rock music on young children cannot be ignored since this is the type of popular music that dominates the environment of most

young children. The constant influence of electronic media is in large part the only nurturing children receive in the developmental music aptitude stage.

Gouzouasis also believes that all forms of accompaniment and a capella singing should be incorporated into elementary classes. The PMMA pretest scores and the PMMA posttest scores, divided into high and low groups of audiation abilities, revealed that students with high music aptitude, as measured by PMMA, scored higher on singing tests than did students with low scores, as measured by PMMA. Explanations for high and low singing scores are appropriate for this study. Physical factors other than audiation skills will inevitably control a child's ability to sing in tune with the proper rhythms. Gouzouasis's final comments on the definitions of conception and perception are concise and rooted in Piagetian theory, but are not universally known by music educators. In the context of singing a song, his definitions are very intuitive and appropriate to the study.

The Estrella Study

The purpose of the Estrella study (1992) was to investigate the effects of training and practice on audiation as reflected in music aptitude scores. In his thesis, Estrella challenges the diagnostic validity of the *Advanced Measures of Music Audiation* (AMMA; Gordon, 1990). AMMA is a short music aptitude test designed for first year college music students and high school seniors. While Gordon believes that AMMA can be used to measure the dimensions of *Tonal* and *Rhythm* as well as provide a composite score, he reports high correlations that range from .72 to .78 between tonal dimension and rhythm dimension scores. It is because of those high correlations that Estrella questions the diagnostic value of AMMA. Estrella believes that AMMA only gives music educators a normative evaluation of their students' general audiation ability, not a diagnostic evaluation of student strengths and weaknesses in audiation ability

relative to the tonal and rhythm dimensions of music learning (p.7).

The test consists of 30 pairs of short musical phrases performed on a Yamaha DX-7 synthesizer and recorded on audio cassette. To take this test students must fill in a space on an answer sheet, indicating whether two short musical phrases sound the same, whether they sound different because of tonal change, or whether they sound different because of a rhythm change. The published predictive performance validities for the *Tonal* , *Rhythm* , and *Composite* scores are .77, .75, and .82, respectively (Gordon, 1990, p.11). As reported in the test manual, split-halves and test-retest reliability coefficients range from .80 to .89 (1989).

The problems that Estrella addresses are, first to determine whether pretest-posttest *Tonal* and *Rhythm* difference scores on AMMA for students who receive training and practice in tonal audiation differ from the pretest-posttest *Tonal* and *Rhythm* scores on AMMA of students who do not receive training and practice in tonal audiation. His second problem was to determine the degree to which pretest-posttest difference scores for harmonic function identification and tonal pattern discrimination predict AMMA *Tonal* scores for those students who receive training and practice in tonal audiation.

A total sample of 76 students participated in this study. There were 37 undergraduate music and non-music majors from Temple University (Philadelphia, PA) and 39 seniors from Matthew McNair Senior Secondary School (Richmond, British Columbia). The British Columbia students ranged in age from 17-18 and were selected from band, choir, and music composition classes. The Temple University students ranged in age from 18-21 and were selected from music and non-music major education classes at the university. Random sampling was not possible due to the length of the experimental treatment and because of a shortage of volunteers. Nineteen volunteers from

McNair Senior Secondary (HS-E) were considered as experimental group one and 17 volunteers (ten music majors and seven non-music majors) from Temple University (TU-E) were considered as experimental group two. The two control groups were comprised of another 20 volunteers from McNair Senior Secondary (HS-C, control group one) and another 20 volunteers (10 majors and 10 non-music majors) from Temple University (TU-C, control group two).

The experimental groups completed two computer based pretests before proceeding with the computer assisted lessons. The pretests consisted of a 30 item test of tonal pattern discrimination (TPD) and a 30 item test of tonal function identification (TFI). The computer program designed by Estrella to teach tonal pattern discrimination and tonal function identification was found to be effective in a pilot study (Estrella, 1990). A mean difference of 1.4 points for tonal function discrimination and a mean difference of 0.8 for tonal pattern discrimination ($p<.05$) was statistically significance. Despite statistical significance, Estrella decided not to use the scores as part of his criterion measures since the TPD and TFI had not been investigated for reliability.

The experimental group completed 26, thirty-minute computer assisted lessons in 13 weeks. Following these lessons, the experimental groups completed the two computer based posttests. Use of a computer was believed to eliminate the possibility of instructor bias in the administration of the treatment.

In the computer assisted music training program, the computer uses a synthesized clarinet timbre to demonstrate tonal patterns in a variety of tonalities and functions without tonal syllables. To demonstrate tonal patterns in a variety of tonalities and functions, the computer program uses digitized audio samples of the researcher's voice singing patterns with tonal syllables. Students are taught to identify tonal pattern functions, to determine if two series

of tonal patterns are the same or different, and to determine whether two musical examples played with a clarinet timbre are the same or different in at least one or more pitches.

Attrition reduced the TU-E group to 11 students; the HS-E and HS-C groups were also reduced to fifteen students each. Academic pressure for both the university and high school students accounted for the attrition. The HS-E group also suffered from apathy.

All students were administered AMMA as a pretest prior to instruction. The control group received their normal music instruction during this period. All students in both experimental and control groups were then administered AMMA *(Tonal* and *Rhythm)* as a posttest.

	University	N	High School	N
Experimental Groups	*AMMA Tonal* difference scores *AMMA Rhythm* difference scores	11	*AMMA Tonal* difference scores *AMMA Rhythm* difference scores	15
Control Groups	*AMMA Tonal* difference scores *AMMA Rhythm* difference scores	20	*AMMA Tonal* difference scores *AMMA Rhythm* difference scores	15

Figure 1. Experimental Treatments by Sites, Two by Two Design (Estrella, 1992)

Two, two way analyses of variance (ANOVA) were used to determine the main effects for site (high school vs. university), treatment (experimental vs, control), and an interaction effect between site and assignment ($p<.05$) for either or both AMMA *Tonal* and *Rhythm* pretest scores (p. 24). No significant differences were found for any of these factors. A two-way ANOVA was used to determine if there were main effects for site, treatment, and an interaction effect between site and treatment (p. 27). A significant main effect for treatment was found.

Two, two-tailed t-tests of mean differences for correlated samples were

calculated for each of the four groups to determine posttest-pretest differences on each measure, tonal and rhythm ($p<.05$) The analyses indicated that AMMA *Tonal* scores decreased significantly for the university control group (TU-C) and increased significantly for the high school experimental group (HS-E). Scores for the university experimental group (TU-E) and the high school control group (HS-C) changed insignificantly. Estrella acknowledges that the use of multiple t-tests increased the probability of a Type I error. Lack of built-in measures to control the emotional and motivational climates of the two sites may have been, in Estrella's opinion, the major contributor to the significant differences in university control group (TU-C) and the high school experimental group (HS-E) scores.

AMMA *Rhythm* posttest-pretest scores for the four groups provided conflicting information. Two-tailed t-tests revealed no significant differences for the high school experimental group (HS-E) and the high school control group (HS-C). Two-tailed t-test analysis showed that the mean posttest score of university experimental group (TU-E) was significantly higher than that group's pretest score, while the university control group (TU-C) showed no difference on AMMA *Rhythm* subtest scores (p. 46).

Based on the results, Estrella believes that both tonal and rhythm audiation, as measured by AMMA, can be altered through specialized training in tonal audiation. Rhythm audiation can also be improved by enrollment in a university music program. Estrella's recommendations include a suggestion to redesign AMMA to include dimensions used in Gordon's MAP model, eliminate the control group, use only one location, and use only the AMMA *Composite* scores as a criterion measure.

A final, controversial conclusion by Estrella is that audiation for students aged 17 to 21 is unidimensional on a Gestalt level, which conflicts with

Gordon's position and that of many other music psychologists, who believe that there are at least seven dimensions of audiation. In the MAP manual, where the word "aptitude" was used instead of the contemporary term "audiation," Gordon narrowed the number of audiation dimensions to seven. That is because these are the most reliable to measure, although he was able to identify in research over 50, highly fluid dimensions of audiation. This researcher believes that audiation, or aptitude, is comprised of many dimensions and that we all have strengths and weaknesses in these dimensions which contribute to our own unique way of creating and enjoying music.

Estrella's study uses a quasi-experimental design to research Gordon's new standardized test for audiation. He questions whether AMMA is a useful instrument for predicting music aptitude or potential, and poses some critical questions pertaining to the value of AMMA as a diagnostic tool for schools and universities. In the AMMA manual, Gordon explains that despite a high correlation between *Tonal* and *Rhythm* scores, a significant portion of the variance for each test is still unique. That statement contradicts Estrella's null hypothesis that the two dimensions are not unique for students aged 17 to 21. Most people, including this researcher, would disagree with Estrella. It seems reasonable that if students at a young age, e.g., before age nine, show strength in either tonal or rhythm aptitude, the same students should continue to exhibit these strengths later in their high school or university years.

The research design used in this study was well defined and clearly written. Replication is possible if a researcher can obtain a copy of the unpublished computer training program. It is unfortunate, however, that a true experimental design was not possible. The use of only student volunteers threatened the internal validity of the study. Despite the use of only volunteers, or perhaps due to the use of volunteers only, the attrition level seems relatively high, especially

for the high school experimental group (HS-E) which decreased 42 percent from 19 to 11 subjects. Estrella did not address the issue of whether boredom or apathy with the computer training program had an effect on student attitudes and their motivation to complete the study. Based on teaching experience, a different approach in the presentation of the program is necessary to sustain the level of interest for both the experimental and the control groups.

Estrella's data analysis for AMMA *Tonal* posttest-pretest score differences determined that there was a significant improvement for the university experimental group (TU-E). Considering the treatment involved training in tonal audiation which was similar to those discrimination abilities tested in AMMA, and the fact that some of the students in the TU-E group were music majors who were exposed to aural skills classes in their university curriculum, these results have little practical significance. Any training effect that resulted from classroom music instruction and ensemble performance experiences could bias the posttest scores in favour of the experimental group. Since none of the students involved in this study received any specific training in audiation prior to the study, the computer training program should have affected latent music abilities.

The use of two sites did offer potential problems, but the researcher believes that the advantages outweighed the disadvantages. No researcher interference was possible for the high school, which allowed the computer training program to stand on its own merit. Computer assisted instruction has some very distinct disadvantages when compared to teacher guided instruction. These include a lack of motivation, boredom, and a lack of accountability. Some students are not visual learners while others need to be motivated through active group participation and aural-oral reinforcement. Pretest and posttest scores for the computer training in tonal skills would have been interesting to analyze and

compare, because of the similarities between the AMMA test questions and the training exercises. Moreover, although the results were not published one may wonder whether the pretest-posttest results of music achievement on the computer program may have contradicted the aptitudinal findings of the researcher.

Estrella's conclusions, based on statistically significant data, are appropriately related to his problems. However, the use of multiple t-tests increases the probability of a Type I error. Also, if Estrella had used an ANOVA, he would have been able to test the differences between all groups and been able to make more accurate probability statements.

The recommendations in this study, although well documented, seem weak. If a researcher used only one test site, he or she might be able to use a two dimensional ANOVA (treatment by levels), where the levels are divided into high and low AMMA pretest scores, and the treatment is administered with a computer program. In that scenario, either a control group or second experimental group might be used. The second experimental group could receive audiation training from the teacher in a traditional classroom environment. The two dimensions could produce more precise data on how well students who possess high or low levels of audiation ability might improve with tonal training over a 13 week period. More research is needed on this topic since the diagnostic value of AMMA is still unproven.

Summary

All of the researchers cited in this review attempted to affect a change in the aptitude scores of students through the use of various instruction and training techniques. Flohr, Jessup, Levinowitz, and Gouzouasis believe that developmental aptitude scores of students through age nine can be influenced, both normatively and idiographically, by various types of instruction. Estrella

was able to affect a change in stabilized aptitude scores for some university and high school students through the use of a computer training and practice program. Based on those conflicting results, the researcher wishes to further investigate that paradigm through the use of computer assisted instruction.

CHAPTER THREE

DESIGN AND ANALYSIS

Sample

Forty-eight Grade 11 and 12 students from Matthew McNair Senior Secondary school in Richmond, British Columbia participated in this study. These students, whose ages ranged from 16 through 19, were enrolled in scheduled school music classes. Matthew McNair Senior Secondary School is a middle class school located in a suburb of Vancouver, British Columbia with an enrollment of approximately 1100 students. Over 25 percent of the school population is made up of ESL (English as a Second Language) students.

Neither group of students received any specialized audiation training prior to the study. Members of the experimental group (T1) were chosen from an intact music composition class. This class represents a sample of 24 students of various music abilities and experience. Before enrolling in this class, students were instructed that completion of the computer program was a necessary partial fullfilment of the course objectives.

Members of the control group (T2) were randomly selected from the remaining students enrolled in music classes. Twenty-four students were randomly selected from a population of 134. A random number table was used for the selection process.

Procedures

Before commencement with the instructional program, all students were administered the *Tonal* subtest of *Advanced Measures of Music Audiation* (AMMA; Gordon, 1989). AMMA is designed for use primarily with university students and consists of 30 pairs of musical phrases. These musical phrases

are recorded on an audio cassette and are performed by a synthesizer. As stated in the oral instructions, the second phrase of each pair may be the same as the first, different because of one or more tonal changes, or different because of one or more rhythm changes. A time of four seconds is given to audiate the two phrases. Students must decide if the two musical phrases are the same or different. If the second phrase is different, the student must choose whether the difference is because of a tonal or rhythm change. Four different answer masks (T1, T2, R1, R2) allow for tonal, rhythm, and total scores to be calculated for this 20 minute test. In the standardization sample of high school students as reported in the test manual, split-halves reliability for the *Tonal* subtest was calculated at .81.

During the 13 weeks of computerized instruction, students in the treatment group completed the *Tonal Syntax Tutorial* (TST, Estrella, 1990), a music training and practice program. A complete printed version of the *Tonal Syntax Tutorial* is provided in Appendix A. Appendix B provides an outline and complete instructions for training procedures. Both appendices are self-explanatory in content and procedures. The training and practice section of this computer program is composed of 26 computer assisted lessons designed to be completed at a rate of two a week. Each lesson takes approximately 20 minutes to complete. Two Macintosh LC computers, placed in individual sound proof practice rooms, were dedicated for student use in this study. The computers were available for students at any time during the school day. Also, students could choose to complete their tutorials during scheduled music class time.

A digitized audio sample of a male voice was used for tonal syllable patterns and a synthesized clarinet timbre was used for the patterns without tonal syllables. In the tutorial, students were given instruction on tonal pattern

identification and function as well as training in determining whether two extended multifunction tonal patterns are the same or different. All of the response data were automatically evaluated and stored by the computer software. In the tutorial, the length of the tonal patterns was gradually increased from six pitches to fifteen pitches in order to aid in the development of tonal memory (Estrella, 1990). The examples performed by the synthesized clarinet are similar to those found in AMMA. Major, minor, mixolydian, dorian, phrygian, and lydian tonalities are all taught and evaluated in the computer software instruction program.

At the end of the 13 weeks of instruction, all of the students were readministered the *Tonal* subtest of AMMA.

Analysis

The null hypothesis is as follows.

H_0: There will be no difference between pretest and posttest tonal aptitude scores, as measured by AMMA, for students who possess high and low levels of tonal audiation ability and who received specialized audiation training, and students who possess high and low levels of tonal audiation ability who received no specialized audiation training.

To determine the effects of computer instruction on the tonal aptitude of the students, the AMMA pretest and posttest scores were organized into a multidimensional design. In the two factor analysis, the A dimension is considered as type of instruction (computer assisted and control group), and the B dimension is considered levels of music aptitude (high and low audiation abilities). In the statistical hypothesis, the dependent variables were the pretest and posttest AMMA *Tonal* scores. Means and standard deviations of the AMMA *Tonal* test scores were calculated. A covariate analysis of AMMA scores was calculated to increase precision and to control for statistical error in the effect of differences in tonal aptitude. To determine differences between the

pretest and posttest scores, a MANOVA was employed. Main effects, interaction effects, and simple main effects were tested at the .05 level of significance.

CHAPTER FOUR

RESULTS AND INTERPRETATION

All 48 Grade 11 and 12 students at Matthew McNair Senior Secondary school completed the13 week study.

For the following reasons, the researcher chose to accept the published test reliability: 1) the reliability of the *Tonal* subtest, as published in the test manual, is high (Gordon, 1989; r=.81); 2) financial constraints prohibited the researcher from calculating a split-halves reliability coefficient; 3) standardized test instructions were adhered to; 4) the researcher is experienced and has administered the test on other occasions; and 5) there were no systematic errors, constant errors, or dubious testing effects that occured during test administration. Means and standard deviations for AMMA *Tonal* pretest scores are presented in Table 1. Also, means and standard deviations for AMMA *Tonal* posttest scores are presented in Table 2.

Table 1. AMMA *Tonal* Pretest Means and Standard Deviations

Variable	Experimental		Control		Total	
High	Mean	**26.75**	Mean	**27.45**	Mean	**27.13**
	SD	1.76	SD	2.94		
	n	12	n	11	n	23
Low	Mean	**22.25**	Mean	**21.92**	Mean	**22.08**
	SD	1.91	SD	2.13		
	n	12	n	13	n	25
Totals	Mean	**24.50**	Mean	**24.45**		
	n	24	n	24		

Table 2. AMMA *Tonal* Posttest Means and Standard Deviations

Variable	Experimental		Control		Total	
High	Mean	**28.00**	Mean	**30.09**	Mean	**29.00**
	SD	5.97	SD	5.71		
	n	12	n	11	n	23
Low	Mean	**26.75**	Mean	**25.07**	Mean	**25.88**
	SD	6.45	SD	3.66		
	n	12	n	13	n	25
Totals	Mean	**24.37**	Mean	**27.37**		
	n	24	n	24		

The MANOVA summary for the AMMA pretest and posttest scores is presented in Table 3. A covariate analysis of AMMA pretest scores revealed that a significant difference existed ($p<.006$) between the students who possessed high tonal aptitude and students who possessed low tonal aptitude before the study began.

Table 3. MANOVA Summary Table for the AMMA Pretest and Posttest Scores

Source	Sum of Squares	*df*	Mean Square	F	*p*
Treatment (A effect)	.203	1	.203	.008	.931
Aptitude (B effect)	116.813	1	116.813	4.409	.042
Trt.x Apt. (AxB effects)	21.225	1	21.225	.801	.376
Explained	357.944	4	89.486	3.377	.017
Residual	1139.306	43	26.495		
Total	1497.250	47	31.856		

As can be seen in Table 3, there was no significant difference between the treatment and control group tonal aptitude scores. Even though there was a significant difference between levels of aptitude, the difference between the students who possess high tonal aptitude and students who possess low tonal aptitude is not a real difference because the difference actually existed in the pretest scores (significant at the $p<.006$ level with AMMA pretest scores used a covariate) before the study began.

CHAPTER FIVE

SUMMARY AND CONCLUSIONS

Summary

With the intent of learning more about the process of assessing music ability, the purpose of this study is to investigate the effects of music learning on music aptitude scores and tonal skills. The problem of this study is to determine if there is a difference between pretest-posttest tonal aptitude scores, as measured by *Advanced Measures of Music Audiation* (AMMA, 1989), for students who possess high and low levels of tonal audiation ability and who received either specialized audiation training or no audiation training.

Forty-eight Grade 11 and 12 students from Matthew McNair Senior Secondary school in Richmond, British Columbia participated in the study. Students were enrolled in scheduled school music classes and their ages ranged from 16 through 19. Matthew McNair Senior Secondary school is a middle class school located in a suburb of Vancouver, British Columbia with an enrollment of approximately 1100 students. Over 25 percent of the school population is made up of ESL (English as a Second Language) students.

Neither group of students had received any specialized audiation training prior to the pretest of AMMA. Students who participated in the Estrella study in 1991 had graduated and were not an influence on this study. Attrition was not a factor in this study because of the thoughtful, yet demanding approach of the instructor. Also, fortunately, none of the students moved to another geographical location during the study. Members of the experimental group (T1) were chosen from an intact music composition class. This class represented a sample of 24 students of various exhibited abilities and

experience. Before students enrolled in this class during February 1992, they were instructed that they would complete the computer program as a partial fulfillment of the course objectives. Members of the control group (T2) were randomly selected, through the use of random number table, from a group of all students enrolled in music classes. Twenty-four students were randomly selected from a population of 140.

AMMA was administered to all of the students as a pretest. *Tonal* subtest scores were used as the criterion measures for this study. The subtest is composed of 30 questions. The experimental treatment consisted of a computer assisted software program, *Tonal Syntax Tutorial* (Estrella, 1991), which provided audiation practice for high school and college students. The control group only received regular music class instruction.

The training and practice section of the computer program is comprised of 26 computer assisted lessons designed to be completed at a rate of two per week. Each lesson takes approximately 20 minutes to complete. Two Macintosh LC computers were dedicated to this study and placed in individual sound proof practice rooms. The computers were available for students at any time during the school day. Most students chose to complete their lessons during scheduled class time. Students in the experimental group were given computer instruction on tonal pattern identification and function as well as training in determining whether two tonal patterns are the same or different. The length of the patterns is gradually increased from six pitches to fifteen pitches in order to aid in the development of tonal memory. The examples performed by the synthesized clarinet are similar to those found in AMMA.

After 13 weeks of computer or non-computer instruction, the *Tonal* subtest of AMMA was readministered to all students. The researcher found no significant difference between the treatment and control group tonal aptitude scores,

however, there was a significant difference between levels of aptitude. It was interpreted that the difference between the students who possessed high tonal aptitude and students who possess low tonal aptitude was not a real difference because the difference in student tonal aptitude levels actually existed before the study began.

Conclusions

Based on the results, the researcher believes that senior level high school students possess stabilized tonal aptitude. These findings are consistent with other research studies (Tarrell, 1965; Fosha, 1960) of music aptitude as defined by audiation. Because of the findings, this researcher believes that AMMA can be a useful instrument in assessing high school student music abilities. Also, based on our present knowledge of computer assisted instruction, that type of instruction alone is not sufficient to affect a change in music audiation ability.

Recommendations

Based on observation, computers are a useful teaching tool for drill and practice, but they are not a substitute for either the instruction or musicianship that can be provided by a music teacher. Moreover, without an instructor present, the motivation of some students may be adversely affected by computer instruction, in that they are unable to focus consistently on music tasks presented by computers. Perhaps future improvements in technology, on both a software and hardware level, will allow students to work more interactively and provide better delivery of music instruction.

Senior Secondary music teachers should feel confident to use AMMA as an assessment instrument. Through the use of AMMA, music teachers will be better prepared to determine the achievement potential of individual students and to base future instruction on various music abilities.

In the future, researchers may wish to study the effect of different types of tonal training on students who possess low tonal aptitude to determine if specialized training will significantly increase their audiation levels. With directed study, students who possess low tonal aptitude may have a greater opportunity to better share in the joy of listening to and creating music. If motivation is considered as a primary factor in the music learning process, longitudinal validity studies may be undertaken to study the effect of instruction over extended periods of time on stabilized music aptitude and music achievement test scores.

BIBLIOGRAPHY

Azzara, C. (1991). Audiation, improvisation, and music learning theory. *The Quarterly, Volume II, Numbers 1&2.* 106-109.

Colwell, R. (1979). *Silver Burdett Music Competency Tests.* Morristown, N J: Silver Burdett Co.

Colwell, R. and Abrahams, F. (1991). Edwin Gordon's contribution: an appraisal. *The Quarterly, Volume II, Numbers 1&2.* 19-36.

Estrella, S. G. (1990). A pilot study to determine the effectiveness of a computer based aural skills training program. Unpublished paper, Temple University, Philadelphia, PA.

Estrella, S. G. (1992). A study to determine the effects of training and practice on advanced measures of audiation. Doctoral Dissertation, Temple University, Philadelphia, PA .

Estrella, S. G. (1991). *Tonal Syntax Tutorial.* Unpublished computer application, Temple University, Philadelphia, PA .

Ferrell, J. (1961). A validity investigation of the *Drake Music Aptitude Tests.* Doctoral Dissertation, Iowa City: The University of Iowa.

Fosha, R. V. (1960). A study of the concurrent validity of the *Musical Aptitude Profile.* Doctoral Dissertation, Iowa City: The University of Iowa.

Flohr, J. W. (1981). Short-term music instruction and young children's developmental music aptitude." *Journal of Research in Music Education, 29,* 219-223. Chicago: G.I.A. Publications.

Gesell, A. L. (1929). Maturation and infant behavior pattern. *Psychological Review, 36.,* 307-319.

Grunow, R. F. (1991). The evolution of rhythm symbols. *The Quarterly, Volume II, Numbers 1&2.* 97-105.

Gordon, E. (1965). *Music Aptitude Profile.* Boston: Houghton-Mifflin.

Gordon, E. (1970). First year results of a five year longitudinal study of the music achievement of culturally disadvantaged students. *Journal of Research in Music Education, Vol. 18, no.3,* Fall 1970.

Gordon, E. E. (1970). *Iowa Tests of Musical Literacy.* Iowa City: The Bureau of Educational Research and Service of the University of Iowa.

Gordon, E. E. (1971). *The Psychology of Music Teaching.* Englewood Cliffs: Prentice Hall.

Gordon, E. (1979). Developmental aptitude as measured by the *Primary Measures of Music Audiation. Psychology of Music, 7,* 42-29.

Gordon, E. E. (1979). *Primary Measures of Music Audiation.* Chicago: G.I.A. Publications.

Gordon, E. E. (1982). *Intermediate Measures of Music Audiation.* Chicago: G.I.A. Publications.

Gordon, E. E. (1986). *The Nature, Description, Measurement, and Evaluation of Music Aptitudes.* Chicago: G.I.A. Publications, Inc.

Gordon, E. E. (1986). Manual for the *Primary Measures of Music Audiation* and the *Intermediate Measures of Music Audiation..* Chicago: G.I.A. Publications.

Gordon, E. (1989). *Advanced Measures of Music Audiation.* Chicago: G.I.A. Publications, Inc.

Gordon, E. (1990). *Predictive Validity Study of AMMA.* Chicago: G.I.A. Publications, Inc.

Gouzouasis, P. J. (1987). The comparative effects of three types of accompanimemt media on the singing achievement and developmental tonal aptitude of young children. Unpublished master's thesis. Temple University, Philadelphia, PA.

Gouzouasis, P. (1990). An investigation of the comparative effects of two tonal pattern systems and two rhythm pattern systems for playing the guitar. Doctoral Dissertation, Temple University. *Dissertation Abstracts International, 51*: 2672-A.

Heads Music Data Summaries of the Higher Education Arts Data Service (Heads). (1986-87). Prepared for the National Association of Schools of Music.

Hohn, R.L. (1991). An educational psychologist considers the work of Edwin Gordon. *The Quarterly, Volume II, Numbers 1&2.* 10-18.

Holmes, J. (1954). Increased reliabilities, new keys and norms for a modified *Kwalwasser-Dykema Test of Musical Aptitude. Psychology of Music, 85.* 38-56.

Jessup, L. L. (1984). The comparative effects of indirect and direct music teaching upon the developmental music aptitude and music achievement of early primary children. Doctoral Dissertation, Temple University, Philadelphia, PA. *Dissertation Abstracts International, 45:* 1678-A.

Kwalwasser, J. and Dykema, P. (1930). *Kwalwasser-Dykema Music Tests (K-D)*. New York: Carl Fischer.

Levinowitz, L. M. (1985). An experimental study of the comparative effects of singing songs with words and without words on children in kindergarten and first grade. Doctoral Dissertation, Temple University, Philadelphia, PA . *Dissertation Abstracts International, 48,*: 863-A.

McDonald, J. C. (1991). Teaching the recorder. *The Quarterly, Volume II, Numbers 1&2.* 110-117.

Pear, T. H. (1911). The classification of observers as 'musical' and 'unmusical'. *British Journal of Psychology, 4.* 89-94.

Piaget, J. (1952). *The Origins of Intelligence in Children.* (Margaret Cook, trans.). New York: International Universities Press.

Revesz, G., (1953). *Introduction to the Psychology of Music.* (G.I.C. de Courcy, trans.) Norman: University of Oklahoma Press.

Saunders, T.C. (1991). The stages of music audiation. *The Quarterly, Volume II, Numbers 1&2.* 131-137.

Schuler, S. C. (1991). Music learning theory to the music education professional. *The Quarterly, Volume II, Numbers 1&2.* 37-58.

Seashore, C. (1919). *Seashore Measures of Musical Talent.* New York: Columbia Phonograph Company.

Schuter-Dyson, R. and Gabriel, C. (1981). *The Psychology of Music Ability.* London: Metheun.

Silver Burdett Music Series: *Book Two.* (1981). Crook, E., Reimer, B., and Walker, D. S. General Learning Corporation, Morristown, N J.: Silver Burdett Co., 24-38.

Stanton, H. (1922). The inheritance of specific musical capacities. *Psychological Monographs ,31*, 157-204.

Tarrell, V. V. (1965). An investigation of the validity of the *Musical Aptitude Profile. Journal of Research in Music Education, 13.*

Trusheim, W. H. (1991). Audiation and mental imagery. *The Quarterly, Volume II, Numbers 1&2.* 138-147.

Walters, D. L. (1991). Edwin Gordon's music aptitude work. *The Quarterly, Volume II, Numbers 1&2.* 65-72.

Woodruff, L. C. (1983). A predictive validity study of the *Primary Measures of Music Audiation*. Doctoral Dissertation, Temple University, Philadelphia, PA. *Dissertation Abstracts, 45*, 01A.

APPENDIX A

TESTS OF TONAL PATTERN DISCRIMINATION AND TONAL FUNCTION IDENTIFICATION

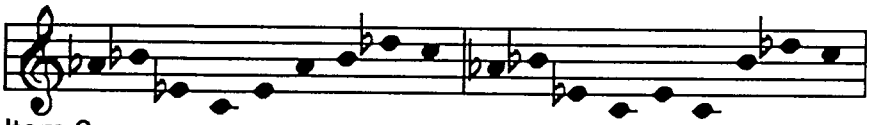
Conventions: Text in *Italics* indicates words displayed by the computer to be read by the student. Text in **Bold** indicates words spoken by the computer using sound samples of the researcher's voice. Text in ***Bold Italics*** indicates words displayed and spoken by the computer to be read and heard by the student. All music notation contained here is for the reader's convenience. The criterion measures contained no music notation component.

Welcome to the Tonal Syntax Tutorial. You will be given two tests to measure your aural skills. Relax and concentrate. Use your intuition and do your best. Begin.

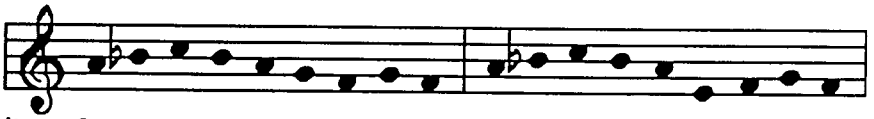
Directions for Test 1
Tonal Pattern Discrimination
I will play two extended tonal patterns on my clarinet. You must determine if the two patterns are the same or different. There will be 30 questions. Good luck. Begin.

TPD questions 1-30 follow. In each item the two extended tonal patterns are separated by a barline.

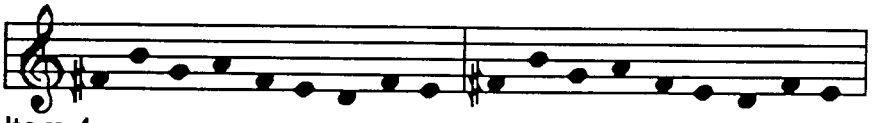
Item 1



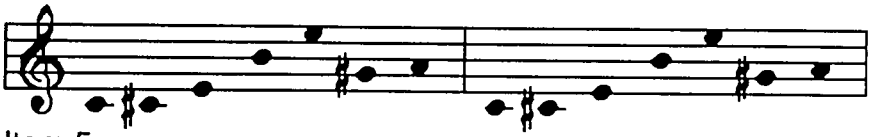
Item 2



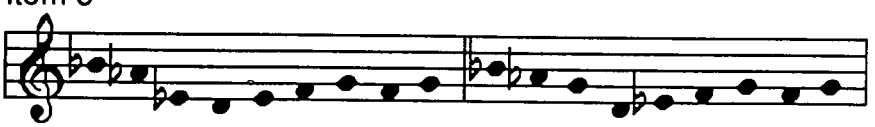
Item 3



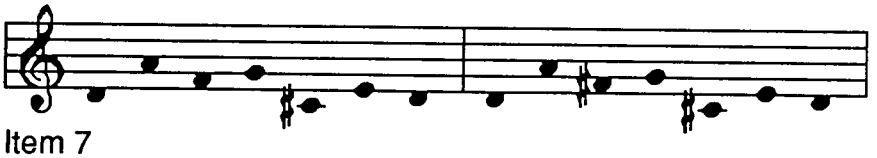
Item 4



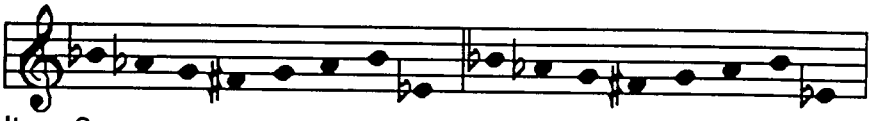
Item 5



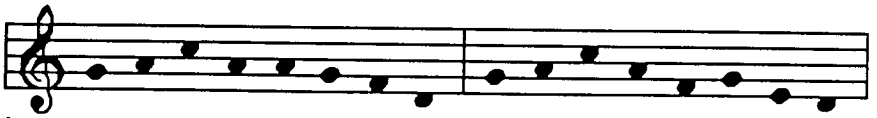
Item 6



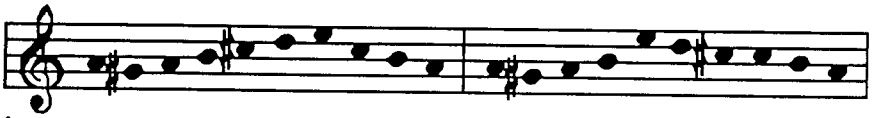
Item 7



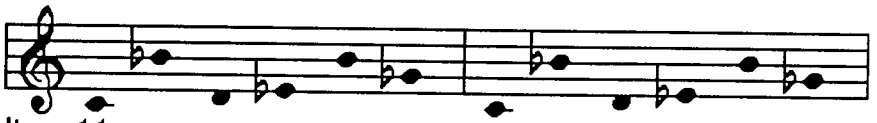
Item 8



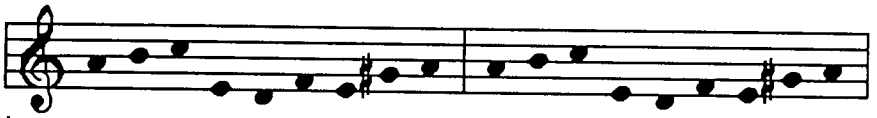
Item 9



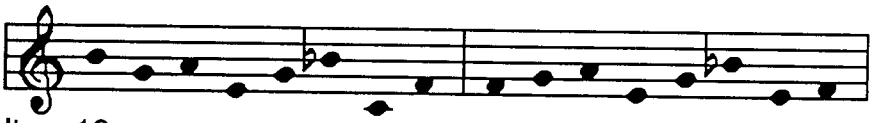
Item 10



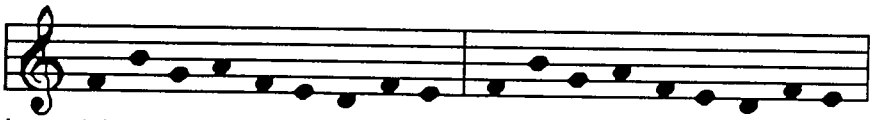
Item 11



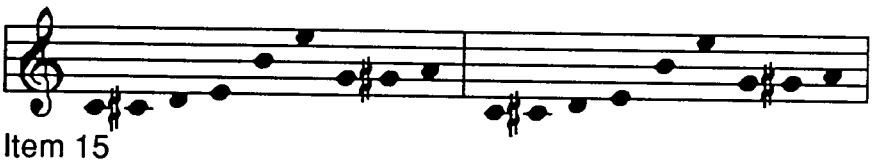
Item 12



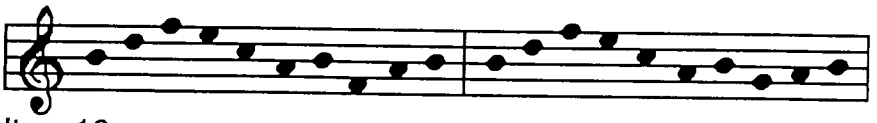
Item 13



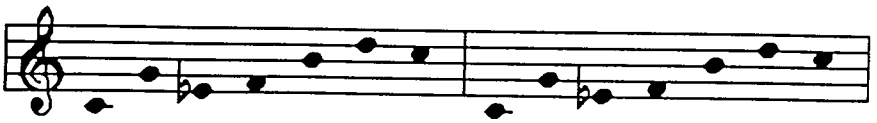
Item 14



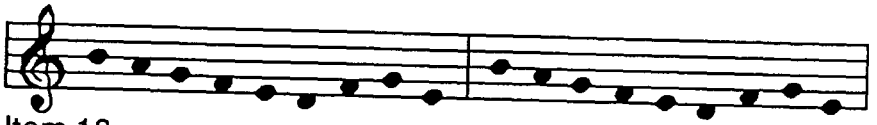
Item 15



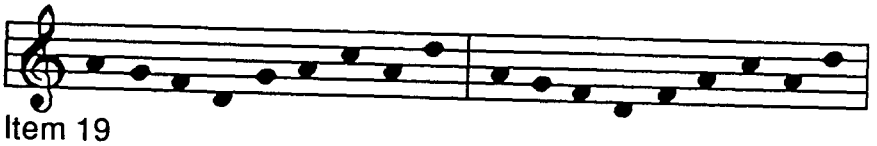
Item 16



Item 17

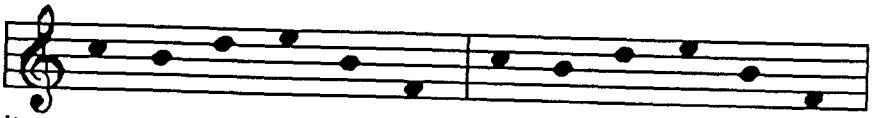


Item 18

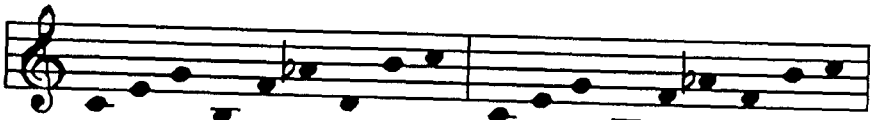


Item 19

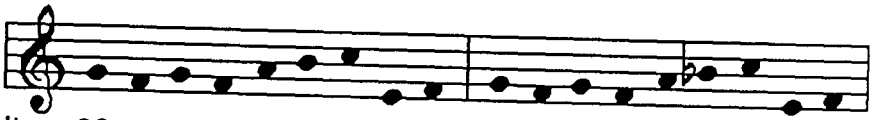
Item 20



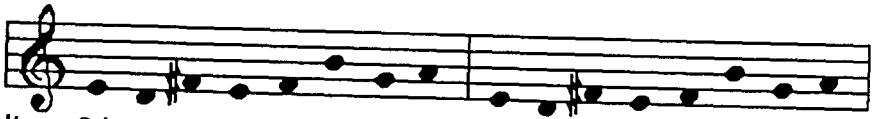
Item 21



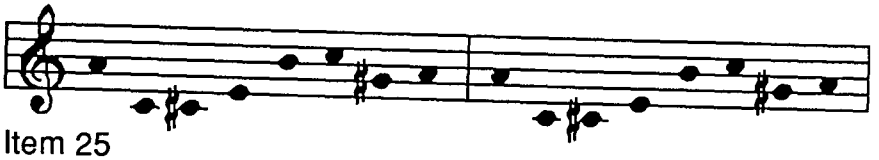
Item 22



Item 23

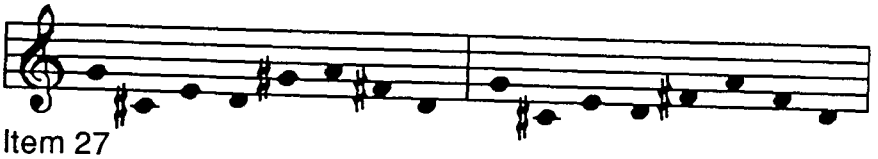


Item 24

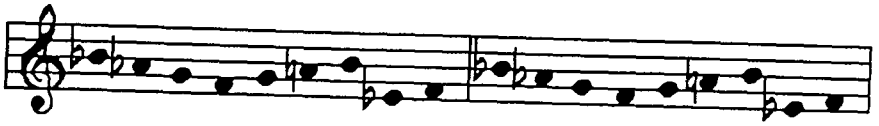


Item 25

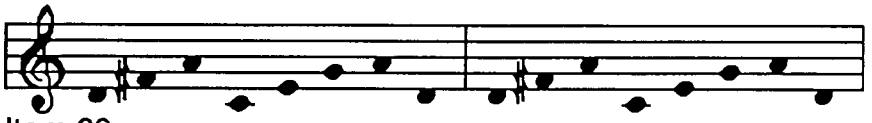
Item 26



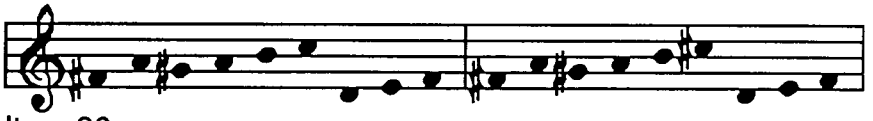
Item 27



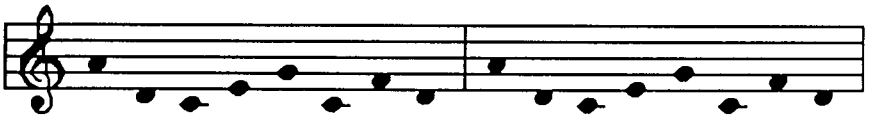
Item 28



Item 29



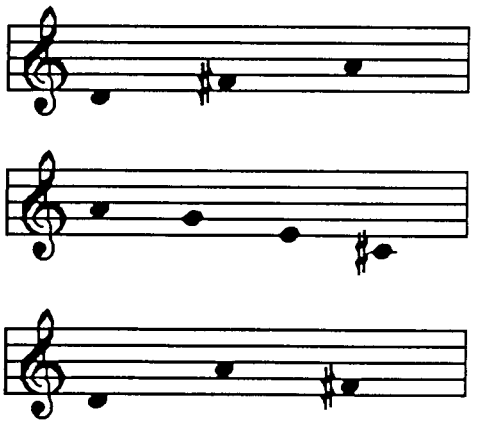
Item 30



Directions for Test 2 Part 1
Tonal Function Identification in Major Tonality
I will play three tonal patterns in major tonality on my clarinet. The first pattern will be tonic, the second pattern will be dominant. Your task is to identify the third pattern as tonic, dominant, or subdominant.

If the third pattern is tonic, then it will contain the same pitches you heard in the first pattern but they might be in a different order. If the third pattern is dominant, then it will contain the same pitches you heard in the second pattern but they might be in a different order. If the third pattern is subdominant, then it will contain a grouping of pitches that is not present in either of the first two patterns.

Listen to this sample question.

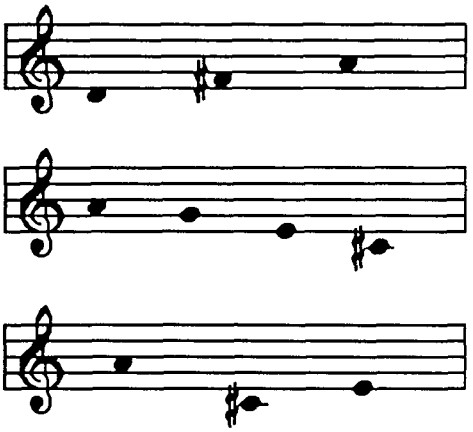


Click on the function of the last pattern
SubDominant Dominant Tonic
If the student answers correctly the computer responds as follows.
***Correct**. The third pattern is tonic, because it contained the same pitches that were present in the first pattern but in a different order.*

If the student answers incorrectly the computer responds as follows.

Incorrect. *The third pattern is tonic, because it contained the same pitches that were present in the first pattern but in a different order.*

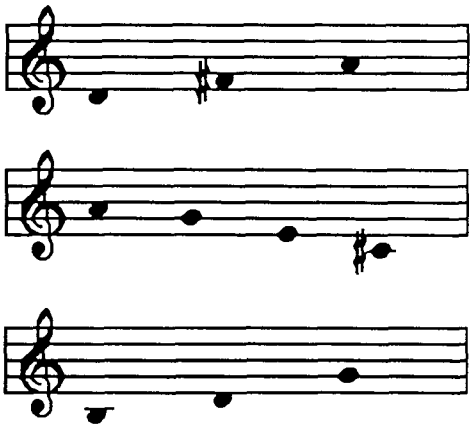
Listen to another sample question.



Click on the function of the last pattern
SubDominant Dominant Tonic
If the student answers correctly.the computer responds as follows.
Correct. *The third pattern is dominant, because it contained a grouping of pitches that were present in the second pattern but in a different order.*

If the student answers incorrectly.the computer responds as follows.
Incorrect. *The third pattern is dominant, because it contained a grouping of pitches that were present in the second pattern but in a different order.*

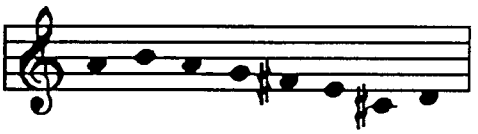
Listen to a final sample question.



Click on the function of the last pattern.
SubDominant Dominant Tonic
If the student answers correctly.the computer responds as follows.
Correct. *The third pattern was subdominant, because it contained a grouping of pitches not present in either of the first two patterns.*

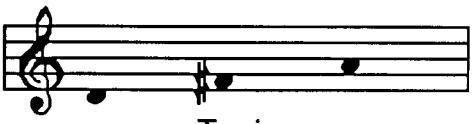
If the student answers incorrectly.the computer responds as follows.
Incorrect. *The third pattern was subdominant, because it contained a grouping of pitches not present in either of the first two patterns.*

There will be 15 questions.
First listen to the preparatory pattern.

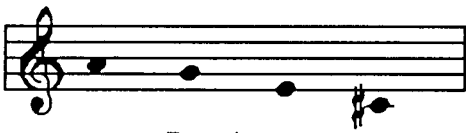


Begin

TFI items 1-15 follow. In each item the first two patterns presented are always the following.



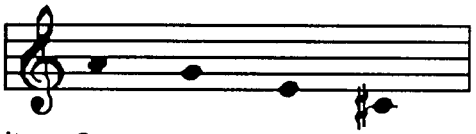
Tonic



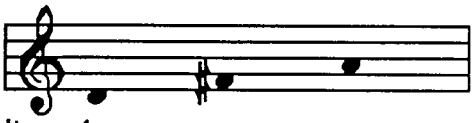
Dominant

The third pattern for each item is listed below
Item 1

Item 2

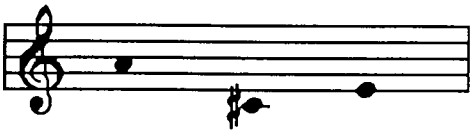


Item 3



Item 4

Item 5



Item 6

Item 7

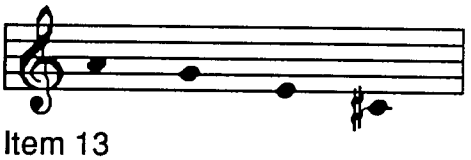
Item 8

Item 9

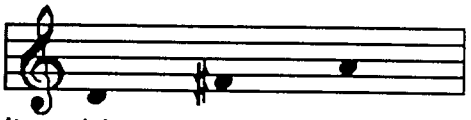
Item 10

Item 11

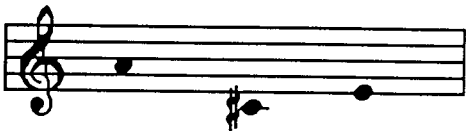
Item 12



Item 13



Item 14

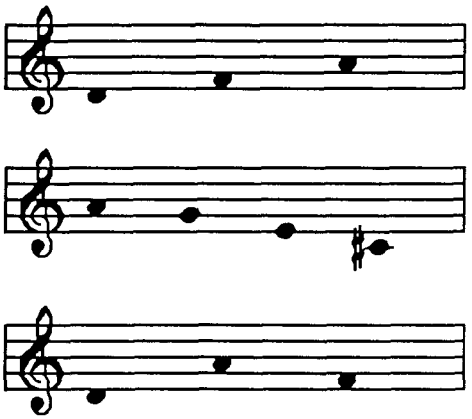


Item 15

Directions for Test 2 Part 2 – Tonal Function Identification in Minor Tonality
I will play three tonal patterns in minor tonality on my clarinet. The first pattern will be tonic, the second pattern will be dominant. Your task is to identify the third pattern as tonic, dominant, or subdominant.

If the third pattern is tonic, then it will contain the same pitches you heard in the first pattern but they might be in a different order. If the third pattern is dominant, then it will contain the same pitches you heard in the second pattern but they might be in a different order. If the third pattern is subdominant, then it will contain a grouping of pitches that is not present in either of the first two patterns.

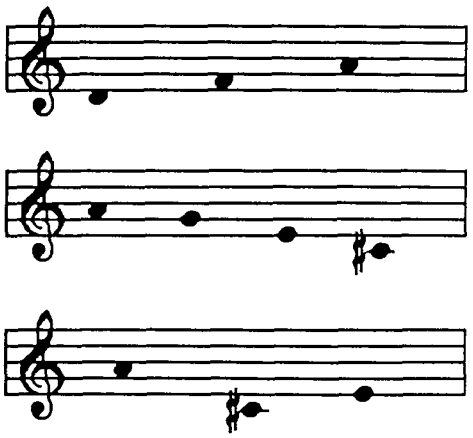
Listen to this sample question.



Click on the function of the last pattern
SubDominant Dominant Tonic
If the student answers correctly the computer responds as follows.
***Correct**. The third pattern is tonic, because it contained the same pitches that were present in the first pattern but in a different order.*

If the student answers incorrectly the computer responds as follows.
***Incorrect**. The third pattern is tonic, because it contained the same pitches that were present in the first pattern but in a different order.*

Listen to another sample question.

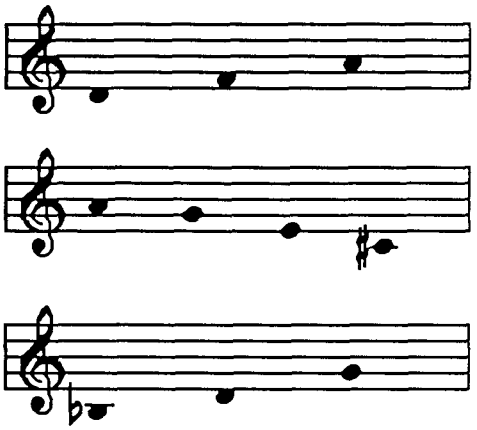


Click on the function of the last pattern
SubDominant Dominant Tonic

If the student answers correctly the computer responds as follows.
Correct. *The third pattern is dominant, because it contained the same pitches that were present in the second pattern but in a different order.*

If the student answers incorrectly the computer responds as follows.
Incorrect. *The third pattern is dominant, because it contained the same pitches that were present in the second pattern but in a different order.*

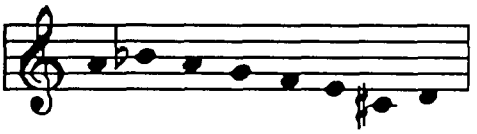
Listen to a final sample question.



Click on the function of the last pattern
SubDominant Dominant Tonic
If the student answers correctly the computer responds as follows.
Correct. *The third pattern was subdominant, because it contained a grouping of pitches not present in either of the first two patterns.*

If the student answers incorrectly the computer responds as follows.
Incorrect. *The third pattern was subdominant, because it contained a grouping of pitches not present in either of the first two patterns.*

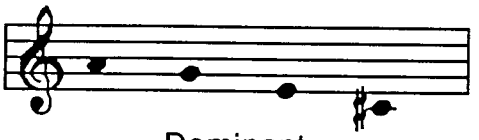
There will be 15 questions.
First listen to the preparatory pattern.



Begin

TFI questions 16-30 follow. In each item the first two patterns presented are always the following.

Tonic



Dominant

The third pattern for each item is listed below.

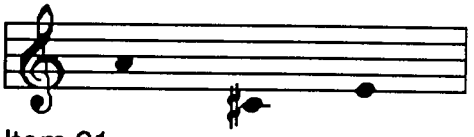
Item 16

Item 17

Item 18

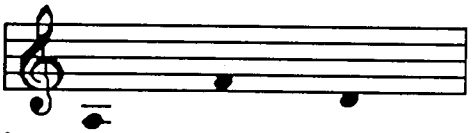
Item 19

Item 20



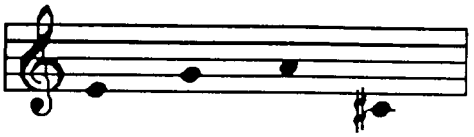
Item 21

Item 22



Item 23

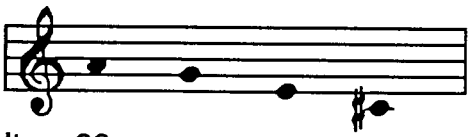
Item 24



Item 25

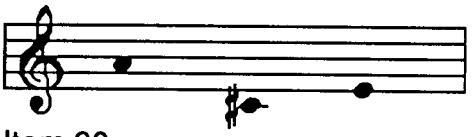
Item 26

Item 27



Item 28

Item 29



Item 30

It has been a pleasure working with you. Goodbye.

APPENDIX B

Procedures for the Training Mode

1. A preparatory pattern in major tonality is delivered, using tonal syllables.
2. The computer informs the student that any combination of "do, mi, and so" is called a tonic pattern.
3. Two tonic patterns are demonstrated, using tonal syllables.
4. The computer informs the student that any combination of "so, fa, re, and ti" is called a dominant pattern.
5. Two dominant patterns are demonstrated, using tonal syllables.
6. The computer informs the student that any combination of "do, fa, and la" is called a subdominant pattern.
7. Two subdominant patterns are demonstrated, using tonal syllables.
8. Each function is demonstrated for review.
9. The computer informs the student that he/she is about to hear three tonal patterns sung with tonal syllables. The first and second patterns are identified as tonic and dominant, respectively. The third pattern is identified by the student as tonic, dominant, or subdominant.
10. If the student answers correctly, another exercise of the same format is presented. If the student answers incorrectly, the computer adds one (1) to the tally of incorrect responses and another exercise of the same format is presented. After ten exercises are presented, the computer presents a number of additional exercises equal to the number of incorrect responses recorded.
11. The computer informs the student that he/she is about to hear three tonal patterns played with a clarinet timbre. The first and second patterns are identified as tonic and dominant, respectively. The third pattern is identified by the student as tonic, dominant, or subdominant.
12. If the student answers correctly, another exercise of the same format is presented. If the student answers incorrectly, the computer adds one (1) to the tally of incorrect responses and another exercise of the same format is presented. After ten exercises are presented in this format, the computer will present a number of additional exercises equal to the number of incorrect responses recorded.
13. The computer explains to the student that sometimes two tonal patterns are almost the same. Two slightly different tonal patterns are played with a synthesized clarinet timbre. The computer then explains why the two patterns are different. The two patterns may be different for a variety of reasons:
    a. the second pattern contains the same tonal functions but with one or more different notes,

b. the second pattern contains a large skip not found in the first pattern,
c. the second pattern involves a different approach to the resting tone,
d. one of the patterns contains a subtonic and the other pattern contains a leading tone,
e. the second pattern is in a tonality different from that of the first pattern, or
f. the second pattern contains a reordering of pitches found in the first pattern.

The two patterns are then played again. Three examples of this type are demonstrated for the student. During the twenty-six instructional sessions, the six conditions were rotated in groups of three so that by the end of training each student experienced each condition four times.

14. The computer informs the student that he/she is about to hear two tonal patterns that are slightly different. The student is then asked to determine why the two patterns are different. If the student answers incorrectly, the computer reveals the correct answer and plays the two patterns again. Five exercises of this type are presented.

Procedures for the Practice Mode

1. A preparatory pattern in major tonality is delivered using a synthesized clarinet timbre.
2. Each tonal function is demonstrated once for review.
3. The computer informs the student that he is about to hear three tonal patterns played with a clarinet timbre. The first and second patterns are identified as tonic and dominant, respectively. The third pattern is identified by the student as tonic, dominant, or subdominant. The specific tonal functions vary depending on the tonality being taught. Table 1 shows the tonal functions to be used for each of the seven tonalities. Twenty items are presented in this format.
4. The computer informs the student that he/she is about to hear a series of two tonal patterns. The first and second patterns may be the same or may differ in one or more pitches. The student must decide if the first and second patterns are the same or different. Twenty items are presented in this format. The first four items are in the same tonality as the tonal patterns in practice procedures 1-3. The tonalities of four of the remaining sixteen items are also in the same tonality as the tonal patterns in practice procedures 1-3. The tonalities of the remaining twelve items were selected according to the distribution of tonalities outlined in Table 2.
5. The computer informs the student that he/she is about to hear two songs played with a clarinet timbre. The first and second songs may be the same or may differ in one or more pitches. The student must decide if the first and second songs are the same or different. Twenty items are presented in this format. The first four items are

in the same tonality as the tonal patterns in practice procedures 1-3. The tonalities of four of the remaining sixteen items are also in the same tonality as the tonal patterns in practice procedures 1-3. The tonalities of the remaining twelve items were selected according to the distribution of tonalities in Table 5.

Table 4. Tonal Functions used in Practice Procedures 1 through 3 for Practice in each of the Seven Tonalities

Major	Minor	Aeolian	Mixolydian	Dorian	Phrygian	Lydian
Tonic	Tonic	Tonic	Tonic	Tonic	Tonic	Tonic
Dominant	Dominant	Subtonic	Subtonic	Subtonic	Subtonic	Dominant
Subdom.	Subdom.	Subdom.	Subdom.	Subdom.	Supertonic	Supertonic

Table 5. Distribution of Exercises in Practice Procedures 4 and 5 for Practice in each of the Seven Tonalities

Major	Minor	Aeolian	Mixolydian	Dorian	Phrygian	Lydian
20 major	8 major	6 major	4 major	3 major	2 major	1 major
	12 minor	6 minor	4 minor	3 minor	2 minor	1 minor
		8 aeolian	4 aeolian	3 aeolian	2 aeolian	2 aeolian
			8 mixolyd.	3 mixolyd.	3 mixolyd.	2 mixolyd.
				8 dorian	3 dorian	2 dorian
					8 phrygian	4 phrygian
						8 lydian

Each of the functions presented in Table 4 serves a different purpose within its relevant tonality. In each tonality the tonic is the most stable harmony because it is an extension of the resting tone. The subtonic or dominant function is the least stable harmony and requires resolution to the tonic. The subdominant and supertonic are important because they contain the characteristic tones of several of the tonalities. By incorporating only these functions into the training program, students were given the opportunity to experience the syntax of each tonality without having to learn the names and sounds of every possible function (Estrella, 1990).