This is to certify that the
thesis entitled

PREDICTIVE ABILITY OF MUSIC APTITUDE AND IMPROVISATION READINESS ON THE QUALITY OF STUDENT IMPROVISATIONS

presented by

DAVID S. TOWNSEND

has been accepted towards fulfillment
of the requirements for the

Master of Music degree in Music Education

Major Professor's Signature

May 2, 2005

Date

PREDICTIVE ABILITY OF MUSIC APTITUDE AND IMPROVISATION READINESS ON THE QUALITY OF STUDENT IMPROVISATIONS

By

David Scott Townsend

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF MUSIC

Department of Music Education

2005

ABSTRACT

PREDICTIVE ABILITY OF MUSIC APTITUDE AND IMPROVISATION READINESS ON THE QUALITY OF STUDENT IMPROVISATIONS

By

David Scott Townsend

The purpose of this research was to investigate the nature of creative thinking in music by examining relationships between music aptitude, improvisation readiness and the quality of student improvisations. High school choir members were given the Advanced Measures of Music Audiation (AMMA), the Harmonic Improvisation Readiness Record (HIRR), and the Rhythm Improvisation Readiness Record (RIRR). They then received instruction in harmonic improvisation during choir for eight weeks. After instruction, HIRR and RIRR were administered once again. The teacher also recorded each student improvising. Independent judges then rated these improvisations using researcher-designed rating-scales. Music aptitude was found to be the best predictor of successful harmonic improvisation. The improvisation readiness records were also highly predictive. Further evidence was established for the nature of HIRR and RIRR as music aptitude tests. The rating-scales were reliable for measuring both convergent and divergent thinking aspects of harmonic improvisation.

ACKNOWLEDGMENTS

I would like to thank Dr. John Kratus for his pioneering research in the field of musical creativity. I would also like to thank him for his valuable insights and helpful guidance throughout this study. I would like to thank Dr. Cynthia Taggart for opening my eyes to the possibilities of teaching creative thinking in music, and for introducing me to and instructing me in the methods of quality research. I would also like to thank Dr. Judy Palac for her beneficial suggestions and careful editing.

I would like to thank my parents, Dr. Rick Townsend and Mrs. Linda Townsend, for introducing me to music, educating me in music, and modeling the kinds of musical creativity that I hope to pass on to my children. Finally, I would like to thank my wife, Judi, for her supportive, understanding commitment to my higher education, and for her continuing love and sacrifice that have made it possible.

TABLE OF CONTENTS

LIST OF TABLES

Chapter One

Introduction

More than 50 years have passed since psychologist J. P. Guilford included aspects of creative thinking in his definition of intelligence (Webster, 1987a). The music education profession has witnessed curriculum development with creative thinking as a primary focus for over 30 years. In the late 1960's, music educators from all levels of teaching met at Manhattan College in Purchase, New York. The outcome of this and other such meetings of the Manhattanville Music Curriculum Project (MMCP) was a music curriculum that included critical, reflective, and creative musical experiences for students (Pogonowski, 2001). Pogonowski notes, "the creative-thinking process was the essence of the MMCP" (p. 26). The inclusion of improvisation and composition in the National Standards for Music Education adopted by the Music Educators National Conference is an evidence of the legacy of MMCP (National Arts Education Associations, 1994).

Creative Thinking

What does creative thinking involve, and how does it apply to music? Webster (1987a) has developed a conceptual model of creative thinking in music that includes aspects of Guilford's work. This model describes the interplay between divergent and convergent thinking. Guilford's concepts of divergent production and convergent production are the basis of this model. Simply stated, divergent production is the result of using information to respond in a variety of ways to a given problem. Convergent production is the result of using information to arrive at a single, desired answer to a problem (Guilford, 1967).

Webster (1987a) describes the thought processes leading to divergent and convergent production as follows:

> Divergent thinking involves the generation of many possible solutions to a given problem–a kind of personal brainstorming. Convergent thinking, on the other hand, involves the weighting of those several possibilities and "converging" on the best possible answer. (p. 165)

Webster (1987a) also describes how divergent and convergent thinking function in the creative process. Individuals use their imagination and understanding of the materials being developed during divergent thinking to generate many possibilities. There comes a point in the creative process, however, when individuals must decide which possibilities they will choose and how they will arrange those possibilities. During this stage in the creative process, convergent thinking is the predominant form of cognition.

Guilford further refined his concept of divergent production. As a result of factor analysis, he identified four unique factors (Guilford & Hoepfner, 1971).

1. *Fluency* is the production of multiple answers from the same given information, in limited time.
2. *Flexibility* is the production of shifts of meaning in response to the same given information.
3. *Elaboration* is the production of detail or complexity of information, above that called for in response to given information.
4. *Originality* is the production of responses rare in the population to which the subject belongs, novel, or remotely associated with the given information.

In the last 25 years, some music educators have described these factors in musical terms and have used them to form objectives for instruction and to evaluate musically creative responses (Auh, 1997; Gorder, 1980; Kratus, 1990; McPherson, 1993; Webster, 1979). Students have created music and their products were evaluated for each of the divergent production factors. Auh, Gorder, Kratus, and McPherson also include musical expression as a fifth dimension of musical creativity. They refer to this dimension with terms like *musical expressiveness*, *music sensitivity*, or *musical quality.*

Another common approach to music curriculum in recent years is the structuring and evaluating of musically creative responses in terms of convergent thinking skills. Music educators evaluate student compositions and improvisations based on the tonality, meter, tonal and rhythm patterns, melodic patterns, pitch range, and expressive devices used in the response (Auh, 1997; Brophy, 1999; Gorder, 1980; Gordon, 2001; Kratus, 1994). Auh, Gorder, and Kratus have also combined divergent and convergent thinking concepts in evaluating musically creative responses.

A recent development in the measurement of creative thinking in music is the application of Amabile's consensual assessment technique for measuring creativity in music (Hickey, 2001). The technique is based on the idea that professionals in a field are able to recognize creativity in that field without breaking it down into highly defined components. When using this technique, judges give creative works ratings based on their own subjective ideas of creativity. Creativity is not divided into convergent or divergent factors. A creative work is evaluated as a whole on a scale ranging from not very creative to highly creative, and the judges' own definition of "creative" serves as the criterion.

Research establishing the reliability of measuring musical creativity in this manner is still in its early stages.

Composition and Improvisation

When musicians hear the words "musical creativity," they may think of composing, improvising, or creatively interpreting music through conducting or performance. Although creative analysis of music or creative listening to music could be included in the list of creative musical activities, most of the research in musical creativity has focused upon composition and improvisation. What are composition and improvisation, to what are they related, and how are they different?

According to the *New Harvard Dictionary of Music*, composition in its broadest form is "the activity of creating a musical work; the work thus created" (Randel, 1986, p. 182). In a study involving relationships between composition and various other factors, Webster (1979) defines composition as "the organization of sounds and silence through symbols yielding a symbolized sound structure that can be aurally realized by others" (p. 229). Kratus (1990, 1994) does not require that music be "symbolized" or "realized by others" to be composition, just that it is repeatable by the composer or other performers in its original form. Kratus (1994) makes a distinction between compositional products that are "fixed, replicable sequences of pitches and durations, and compositional processes [which are] the fluid thoughts and actions of the composer in generating the product" (p. 116). Studies have shown that composition is significantly related to music aptitude, music achievement, informal musical experience, self-concept as musicians, age, and academic achievement (Auh, 1997; Kratus, 1994; Laycock, 1992).

Improvisation is "the creation of music in the course of performance" (Randel, 1986, p. 392). Webster (1987a) broadly defines improvisation as "the transmission of sound structures that are actually conceived by the performer at the time of performance" (p. 162). In one study involving improvisation, Azzara (1993) writes, "improvisation means that an individual has internalized a music vocabulary and is able to understand and express musical ideas spontaneously. Improvisation is to music what speaking is to language" (p. 329). When improvising with melody and harmony, Gordon (1998) observes that improvisation may take three forms.

> First, one may perform a variation of a melody without giving attention to the underlying existent or implied harmony. Second, musicians may perform a melody over a series of harmonic patterns, otherwise called harmonic progressions. Third, musicians may improvise harmonic patterns to an old or new melody. (p. 9)

Research has shown that improvisation is significantly related to music achievement, imitative ability, playing by ear, singing, mental rehearsal, composing, playing from memory, age, figural creativity, IQ, and certain personality traits (Brophy, 1999; Madura, 1996; McPherson, 1995; Swanner, 1985; Webster, 1979). Studies that used sound improvisations to measure creative thinking in music have found no significant relationships between music aptitude and musical creativity (Baltzer, 1990; Hickey & Webster, 1999; Schmidt & Sinor, 1986; Swanner, 1985; Webster, 1987a). However, studies that used tonal, rhythmic, and harmonic improvisations of students have found significant relationships between music aptitude and quality of improvisation (Gordon, 1998, 2001; Jessen, 1991; Webster, 1979).

The *New Harvard Dictionary of Music* (Randel, 1986) contrasts composition and improvisation. "The term [composition] is most often used in opposition to improvisation, implying an activity carried out prior to performance or a work whose features are specified in sufficient detail to retain its essential identity from one performance to another" (p. 182). The specifications of composition being pre-planned and replicable are consistent with music educator's definitions as stated earlier. Kratus (1990) notes the following:

> Improvisation is sometimes referred to as "simultaneous composition," but it is not a type of composition. Instead, composition can be thought of as "reflective improvisation," because time to reflect and change musical ideas is an integral part of the process. (p. 36)

If composition is really an extension of improvisation, then music educators should sequence creative activities to include experiences with improvisation as a readiness for experiences with composition. This concept leads one to wonder whether improvisation is also a readiness for other musical activities.

In a study using improvisation as a teaching tool, Azzara (1993) found that students involved in improvisation activities score higher on instrumental performance achievement measures than students not involved in improvisation activities. Just as speaking language in conversation prepares an individual to better understand written language, improvisation prepares a musician to more fully bring meaning to music notation. Gordon (2003) also discusses outcomes of improvisation instruction:

> When a teacher provides students with the skill to create and to improvise their own music, music becomes the property of the students themselves, and this

should represent the ultimate goal of all teachers. In a word, students who learn to create and to improvise music through audiation [see definition below] become critical listeners, not amateur critics. (p. 40)

Azzara (1993) explains why improvisation has connections to so many other areas of music and musicianship:

> Like an individual's contribution to conversation, improvisation in music is generated from an internal source. Improvisation skill allows students to express musical thoughts and ideas from that internal source, with meaning. When thought of in this light, improvisation is at the heart of musical expression and is fundamental to all types and levels of music instruction and curriculum. Teaching students to improvise promotes their acquisition of higher-order music thinking skills. Asking students to improvise as a part of instrumental music instruction would reasonably increase a student's ability to manipulate mentally the structures of music with purpose and meaning. (p. 330)

Music Aptitude and Audiation

As was mentioned earlier in this chapter, researchers have examined the relationship between musical creativity and music aptitude. Music aptitude is "a measure of a student's potential to learn music" (Gordon, 2003, p.41). Gordon makes a distinction between music aptitude and music achievement. When music educators seek to measure what students have already learned in music instead of their potential to learn music, they are dealing with music achievement. Through years of research and teaching music to children, Gordon has developed music aptitude tests for students in pre-kindergarten through college. These tests measure students' tonal aptitude, rhythm aptitude, and in the

case of the *Musical Aptitude Profile*, musical sensitivity. Gordon has also developed a readiness test for improvisation that shows preliminary evidence of being an aptitude test for harmonic and rhythmic improvisation (2001).

Since an individual's potential to audiate music is a part of Gordon's conception of music aptitude, an understanding of the term "audiation" is necessary. When one audiates music, they mentally hear and syntactically understand music whether or not the sound is physically present (Gordon, 1998). In the cases of composition and improvisation, an individual may audiate music that has never before existed as physical sound. Audiation differs from aural perception in that it does not rely on sound being physically present. Audiation differs from inner hearing in that it is not just imagining sound; it requires that musical meaning be added to the sound. It is hearing musical sound in a context of musical syntax (Gordon, 2003). Gordon uses the analogy that "audiation is to music what thought is to language" (1998, p. 9).

References to the concept of audiation are found among composers and improvisers. For example, composers describe the ability to think in sound as being an important part of the composition process (Kratus, 1994). Fine jazz musicians can "hear the changes" in their minds ahead of time and improvise accordingly (Gordon, 1988). Someone who is harmonizing "by ear" audiates the chords and rhythm patterns ahead of time and improvises a harmony part to fit the chord changes and rhythms of the melody.

It seems reasonable that researchers would find a significant, positive relationship between students' potential to audiate and their musically creative responses. Yet, the results of studies to date seem to disagree with each other. It is possible that current tests of music aptitude measure audiation potential as it relates to convergent production, but

not divergent production. As a result, these music aptitude tests do not significantly correlate with musical creativity tests based on divergent production. It is also possible that current music aptitude tests measure audiation potential as it relates to some kinds of musically creative activities, but not to others. It is more likely, however, that the discrepancy is caused by the definition of a "musically creative response" in some tests of creative thinking in music. If a student's music aptitude is measured in terms of tonality and meter, but a student's musical creativity in measured in terms of largely random sound responses, there is little wonder that the two scores do not significantly correlate.

Purpose and Problems

The purpose of this research is to investigate the nature of creative thinking in music by gaining information about the relationships between music aptitude, improvisation readiness and the quality of student improvisations. A more complete understanding of the readiness required for successful harmonic and rhythmic improvisation may enable music educators to teach improvisation more effectively. The specific problems of this study are as follows:

1. To develop a reliable criterion measure that uses both convergent and divergent production factors for evaluating student improvisations.
2. To examine relationships between music aptitude and the quality of student improvisations for evidence of predictive validity.
3. To examine relationships between improvisation readiness and the quality of student improvisations for evidence of predictive validity.

4. To examine relationships between *Harmonic Improvisation Readiness Record* (HIRR) and *Rhythm Improvisation Readiness Record* (RIRR) scores before and after instruction for further validation of the tests as music aptitude tests.

Chapter Two

Measuring Musical Creativity

During the past 20 years there have been numerous studies using improvisation to measure creative thinking and production in music. Some researchers have designed their own criterion measure for evaluating student improvisation. Others have used a previously designed measure. This chapter will describe and analyze representative studies and their criterion measures in terms of their design, effectiveness, and results.

Measures of Creative Thinking in Music (MCTM)

Webster (1987a) designed the MCTM to measure creative thinking in music. The measure is appropriate for 7- to 10-year-old children. Students use a "Nerf" ball to play tone clusters on a piano. They also use a microphone with reverberation and a set of five wooden resonator blocks. Students are given experimentation time with the instruments. They are then asked to create sound improvisations to depict frogs or singing robots. Students are also asked to use the instruments to depict a story and to create a piece of music with a beginning, middle, and end using all of the instruments.

Student responses are evaluated according to the divergent factors of musical extensiveness [fluency], musical flexibility, musical originality, and musical syntax [elaboration]. Webster (1987b) defines these factors as follows:

1. Musical Extensiveness - the actual clock time (in seconds) involved in a musical response.
2. Musical Flexibility - the extent to which a child can freely move from one extreme to another with one of the three musical parameters: low to high, soft to loud, fast to slow.

3. Musical Originality - the extent to which the child manipulates musical phenomena in a unique fashion.
4. Musical Syntax - the extent to which the child manipulates musical phenomena in a logical and inherently musical manner, with attention to the shaping of the whole response and not just a single part. (p. 264)

Musical Extensiveness and Musical Flexibility are measured objectively by watching videotape of the session. Musical Originality and Musical Syntax are measured subjectively using rating scales. Recently, the measure has been revised to include a MIDI keyboard and computer for administering the test and recording data (Hickey & Webster, 1999). The rating scales for Musical Originality and Musical Syntax have been revised as well. Reliability coefficients for the Musical Originality and Musical Syntax factors are reported at .78 and .69 respectively (Webster, 1987b). Content validity and construct validity have been supported by panels of musicians, some factor analysis, and by the use of widely accepted factors of divergent thinking. Test-retest reliability and predictive validity have not yet been studied. Studies using MCTM consistently show little or no relation between creative thinking in music and music aptitude (Baltzer, 1990; Hickey & Webster, 1999; Schmidt & Sinor, 1986; Swanner, 1985; Webster, 1987a, 1987b).

Webster (1987b) uses the fact that music aptitude scores and MCTM scores have no correlation or negative correlation as evidence of inverse concurrent validity for the measure. While this is one interpretation, there are other options that need consideration. In the music aptitude tests correlated with MCTM, responses to sets of rhythm patterns, tonal patterns, or musical phrases determine the music aptitude of an individual.

Therefore, for musical creativity test scores to positively correlate with music aptitude scores, the musical activity and evaluation would probably have to include tonal and rhythm patterns organized into tonality and meter.

The activities in MCTM have students improvise with elements of music such as tempo, dynamics, pitch, timbre, and form. A child does not necessarily need to use the elements of melody, rhythm, or harmony to receive a high score. The child does not even really have to audiate tonally, rhythmically, or harmonically to receive a high score. Student responses need not be in a tonality or meter. In fact, most of the activities are designed to produce sound improvisations devoid of musical syntax. It is not surprising that MCTM scores do not positively correlate with music aptitude scores. Undoubtedly, the test measures some aspects of creative thinking, but whether or not it measures creative thinking in music is debatable.

When Kratus (1994) examined the compositional processes and products of students and compared them to their music aptitude scores, both positive and negative relationships appeared. Music aptitude correlated positively with the process of development and silent reflection. It also correlated positively with compositions that were tonally and metrically cohesive, demonstrated developed rhythm patterns, and were more limited in pitch range. Music aptitude correlated negatively with extended use of exploration and the use of a wide pitch range. Students with less developed audiation skills used more of their time exploring the keyboard and had products that were less musically structured and cohesive.

Measures of Musical Divergent Production (MMDP)

Gorder (1980) designed a criterion measure to assess fluency, flexibility, originality, elaboration, and musical quality in the improvisations of students in school instrumental programs. A listing of musical content categories was developed to help evaluate the improvisation responses. "It included 78 divisions of nine content areas: melodic, rhythmic, pulse/meter, tempo, style, dynamic, timbral, expressive device, and form" (p.36). The divergent production factors were defined in musical terms as follows:

1. Musical Fluency – the production of musical ideas or phrases (content) from given music information.
2. Musical Flexibility – the production of ideas or phrases that are scored for shifts in content character.
3. Musical Elaboration – the production of musical ideas or phrases that are scored for the detail or complexity of content characteristics employed.
4. Musical Originality – the production of musical ideas or phrases that are scored for the use of musical content characteristics rarely used by the population to which the subject belongs.
5. Musical Quality – the production of musical ideas or phrases that appeal to musicians' musical sensitivity. (p. 35)

Students were asked to improvise as many phrases as they could in a limited time. The basis of student improvisations were two different musical motives, an incomplete series of pitches that began and ended on the resting tone, and an incomplete series of pitches with a melodic contour shape included. The items were notated on a musical staff and the students could sing, whistle, or play their improvisations on a familiar instrument.

Responses were tape-recorded. The main study included 40 high school instrumentalists and 40 junior high school instrumentalists.

Reliability coefficients for the divergent thinking factors were moderately high, ranging from .70 to .90. However, the only factor with a range of interjudge reliability above .80 was Musical Fluency (.90-1.00). Content validity was obtained through interviews with professional musicians. Correlations for concurrent validity were examined in relation to teacher ratings of students' musical creativity. Results were mixed. Construct validity for the factors of the MMDP were examined in relation to each other and to other musical tests through factor analysis. MMDP did not show significant correlations with music aptitude, music achievement, age, musical training, or musical experience.

The MMDP is a mix of convergent and divergent thinking factors. Students improvised using musical content and were evaluated on both their convergent and divergent products. Gorder defined Guilford's divergent thinking factors logically in terms of musical responses. Gorder (1980) reported that the scoring procedures were complicated and needed refinement. It is possible that refining the scoring procedures would make the measure even more reliable and help to make conclusions even more meaningful.

Test of Ability to Improvise (TAI)

McPherson (1993) developed TAI to measure the improvisation abilities of high school instrumentalists. One hundred one high school instrumentalists improvised in response to seven different items. Improvisation activities included improvising a closing phrase, improvising a melody to a set rhythm pattern and key, improvising a musical

motive, improvising a melody over an accompaniment, and improvising freely in a style or mood. Aspects of convergent production and divergent production were used to evaluate the five different kinds of student improvisations. Interviews with professional musicians led to a criterion measure that included five-point rating scales using four evaluating criteria:

1. Instrumental Fluency – Hesitant and labored to spontaneous and confident.
2. Musical Syntax – Illogical to logical.
3. Creativity – No uniqueness to marked uniqueness.
4. Musical Quality – Unappealing to appealing. (McPherson, 1995, p. 149)

Each of the evaluating criteria used a five-point rating-scale. Judges listened to the taped improvisations twice. The first time they evaluated the instrumental fluency and musical syntax dimensions. The second time they evaluated the creativity and musical quality dimensions. Student responses were given a score of 1 to 5 for each of the criteria and then combined for a total score. Interjudge reliability correlations were acceptable (.71 to .94) for the 4 evaluative criteria and high (.89 to .97) for the composite scores on each individual item. Convergent and discriminant validity were examined using the multitrait-multimethod approach. Results of the analysis satisfied each of the validity criteria.

TAI demonstrates many positive characteristics. It includes a variety of musical improvisation activities and evaluates them in both convergent and divergent terms. It measures specific musical improvisation traits as well as musical creativity traits. The evaluative criteria are stated simply and accurately enough to be used reliably by different judges. One improvement to TAI could be the use of a continuous rating-scale

with specific criteria instead of just a number range for each of the four evaluating criteria. A rating-scale including specific criteria with each number gives more information about where the student is in each criterion and how to sequence instruction more effectively for further improvement.

Results of the study include positive relationships between improvisation ability as measured by TAI and a variety of factors. Although music aptitude was not examined, McPherson (1993) reports significant positive correlations between TAI scores and performance proficiency, experience in singing, mental rehearsal, playing another instrument, and frequency of improvising. The concept of mental rehearsal is strongly tied to the concept of audiation and could, therefore, suggest that music aptitude and improvisation ability as measured by TAI are related.

Harmonic Improvisation Readiness Record (HIRR) and Rhythm Improvisation Readiness Record (RIRR) Studies

In the last few years, Gordon (2001) has conducted several studies aimed at establishing the concurrent validity of HIRR and RIRR. More recently, he has also examined the predictive validity of HIRR and RIRR for harmonic improvisation. Since the measure seemed to possess characteristics of both music aptitude tests and music achievement tests, post-publication research was also necessary to determine the exact nature of the measure.

In one such study, 33 students in college music theory and music history courses were administered HIRR, RIRR, and the *Advanced Measures of Music Audiation* (AMMA). Then, the students listened individually to six songs. After each song was played twice, students responded with their own song that was like that song, but not an

imitation of that song. Student responses were recorded and evaluated with two different five-point continuous rating-scales. The harmonic rating-scale included the following categories and did not consider keyality or voice quality:

1. Student sings without a sense of tonality.
2. Student sings in the appropriate tonality but without implied harmonic changes beyond tonic function.
3. Student sings in the appropriate tonality with some implied harmonic changes.
4. Student sings in the appropriate tonality with all implied harmonic changes.
5. Student sings in the appropriate tonality with all implied harmonic changes and with musical melodic contour. (p. 22)

The rhythm rating-scale included the following categories and did not consider rapidity of tempo:

1. Student sings without a consistent tempo and/or in the appropriate meter.
2. Student sings in a consistent tempo and in the appropriate meter.
3. Student sings in the appropriate meter with some duration changes coinciding with implied harmonic changes.
4. Student sings in the appropriate meter with all duration changes coinciding with implied harmonic changes and without extending or shortening the melody.
5. Student sings in the appropriate meter with all duration changes coinciding with implied harmonic changes and with musical style. (p. 22)

Results of the study showed that while HIRR and RIRR are somewhat related to AMMA, they are more related to each other. The correlation of the combined HIRR and

RIRR scores with the combined harmonic and rhythm rating-scale scores was .91. The composite AMMA scores only related to the combined harmonic and rhythm rating-scale scores with a correlation of .50. Therefore, even though music aptitude as measured by AMMA has a high relationship to harmonic improvisation, HIRR and RIRR have a much higher relationship with one another and are not measuring the same thing as AMMA.

Another recent study (Gordon, 2001) involving HIRR and RIRR established evidence for predictive validity as well as evidence that the measure is a music aptitude test. Students in specially chosen instrumental music classes from two high schools were administered HIRR, RIRR, and AMMA at the beginning of the semester. The teachers did not know the results of the tests until after the semester was completed. There were 22 students in one class and 21 in the other. Throughout the semester, the students all received instruction in jazz harmonic improvisation. Their teachers used different methodologies and designed their own three-point rating-scales for evaluating student improvisation at the end of the semester. After instruction for the semester, the students were administered HIRR and RIRR again. They also took the IPT.

Test scores from the second administration of HIRR and RIRR were not significantly different from the first administration scores. Results of this study showed that HIRR and RIRR scores were not significantly altered with practice and training. Gordon (2001) reports that HIRR and RIRR are music aptitude tests and that "they account for approximately 55% of the reason or reasons why students learn successfully to improvise harmonically" (p.56). Comparing scores from the first administration of the measure to improvisation performance scores at the end of the semester also established

longitudinal predictive validity for the measure. Since the research used few students in a limited setting, more studies are needed to verify the predictive validity of the measure.

Again, correlations between HIRR and RIRR scores and improvisation performance scores were high. The correlations between HIRR and RIRR were low as well as the correlations between each of them and AMMA. AMMA demonstrated lower predictive validity for harmonic improvisation than either HIRR or RIRR.

The studies described in this chapter leave several questions for one considering the subject of creativity in music. Would creative responses in music correlate more strongly with music aptitude if they were limited to creative responses in tonality and meter? Would creative responses in music correlate more strongly with music aptitude if the measure of improvisatory responses was adequately reliable? Would including descriptions along with each number in a measure of creative musical response increase the information gained by that measure? Would a further study that measures the quality of student improvisations sustain the preliminary evidence of predictive validity for HIRR and RIRR? Finally, would such a study support or detract from the evidence that HIRR and RIRR are music aptitude tests?

Chapter Three

Methodology

The study described in this chapter was designed to examine the previous questions and to help provide answers to those questions. The individuals who partook in this study are described. Then, a general overview of the study is included, followed by a detailed description of the study's teaching procedures and criterion measures.

Subjects

The individuals involved in this study were high school choir students from Freedom Farm Christian School (FFCS). Fifteen students were male and fourteen were female. All were Caucasian, and tended to be from a lower to middle socio-economic status. The students participated in improvisation instruction during choir warm-ups throughout the fourth quarter of their school year. The students were in the ninth through twelfth grades. This study used the test scores of the twenty-nine choir students who completed all activities and testing and who granted me consent to use their test scores and improvisations.

Most of the students from FFCS have received a few years of elementary general music instruction through the seventh grade. Some had been members of the band since the sixth grade. While some had previously experienced harmonic improvisation in band, they had not yet received instruction in vocal improvisation.

The music director at FFCS teaches band, choir, and general music classes K-12 and emphasizes the development of audiation skills throughout the music program. However, many of the students in high school choir this particular year were new to the school and were still developing basic audiation skills and vocal technique.

Design

The students took the *Advanced Measured of Music Audiation* (AMMA) at the beginning of the school year in August 2003. This test determines an individual's ability to audiate tonal and rhythmic aspects of music. The students took the *Harmonic Improvisation Readiness Record* (HIRR) and the *Rhythm Improvisation Readiness Record* (RIRR) in March 2004. These tests determine an individual's ability to hear harmonic changes in music. They test the ability to hear when the harmonies change, and which harmonies are used.

During late March, April, and May, I instructed the students in harmonic improvisation twice a week using methods outlined in the improvisation curriculum, *Creativity in Improvisation* (Azzara, Grunow, & Gordon, 1997). At the end of the school year, all of the students took HIRR and RIRR a second time. I also administered a researcher-designed harmonic improvisation performance test (IPT) after instruction in harmonic improvisation had ceased for the semester. The test measured both convergent and divergent aspects of their harmonic improvisations.

In this study, relationships between students' scores on the AMMA and the IPT were examined for predictive validity. Relationships between the scores of the first administration of HIRR and RIRR and the IPT were also examined for predictive validity. Relationships between the scores from the first administration of HIRR and RIRR and the second administration of HIRR and RIRR were examined to further validate the tests as music aptitude tests. The IPT was examined for reliability.

Overview of the Procedures

At the beginning of the year, I administered AMMA to all music students. Fourth-quarter choir classes began with the administration of HIRR and RIRR. Since AMMA, HIRR, and RIRR are standardized tests, I administered them according to the directions in the test manual. Then, I chose class activities that sequentially lead the students through harmonic improvisation skill development. I followed the general guidelines for improvisation instruction outlined in the improvisation curriculum, *Creativity in Improvisation* (Azzara, Grunow, & Gordon, 1997). Class activities were sequenced in a manner consistent with the IPT.

Five to ten minutes of class time were set aside twice a week to devote to improvisation instruction. The improvisation activities were familiar to some choir students because they had experienced similar instruction in band the last few years. Sixteen of the students were members of both band and choir. However, this was the first time any of the students tried to improvise vocally.

As preparation for improvisation activities, the students learned many songs in a variety of meters and tonalities. First they learned the melody of a song until they could sing it without me. Next they learned the "root melody" for that song. The root melody is the chord roots or the bass part in root position for a song. When students were comfortable singing the melody and root melody separately without my help, I split them into two groups. One group would sing the melody while the other group sang the root melody using a neutral syllable or solfege syllables. The groups were divided in many ways to help them develop the skill of singing different music from those around them.

The students were also taught simple chord progressions in major and minor tonalities to sing as a class. This helped them to gain more vocal independence and develop their harmonic awareness. When they were ready, the students could then substitute an alternate chord tone for pitches of the root melodies they had previously learned. Now I could sing melodies while the class harmonized, or part of the class could sing a melody while the rest of the class harmonized. I indicated which chord the choir should sing by a hand gesture. For example, if the phrase used a I-IV-V^7-I chord progression, I would hold up one finger, then four, then five, then one. After the students were familiar with singing the harmony part, the hand gestures were no longer necessary for that song.

Preparation for improvisation activities also included echoing different tonal and rhythm patterns without, and then with tonal and rhythm solfege. These activities allowed students to gradually build a vocabulary of tonal and rhythm patterns from which to improvise. When the students had learned many tonal and rhythm patterns, they were asked to improvise just rhythms or just pitches. For instance, after the class had echoed many rhythm patterns in duple meter, I might ask volunteers to make up a duple meter rhythm pattern of their own. A similar process took place with improvising pitches. After the class echoed many tonal patterns in major, I might sing the pitches of a tonic pattern and ask if anyone could think of another order for those pitches. Gradually, after students were comfortable improvising rhythm and tonal patterns separately, they were ready to combine them in harmonic improvisation.

Harmonic improvisation activities were designed to lead students into the ability to improvise their own music to given chord progressions. I used five basic levels of

activities to build harmonic improvisation skills. Level one was simply singing the root melody on each macrobeat. In the second level, students improvised rhythm patterns to a root melody using a neutral syllable. The difference between this step and the rhythm improvisations they had previously experienced was that now they were improvising rhythms in a given meter while singing the correct pitches of a root melody. This activity was usually done as a class with each student singing the same pitches as their neighbor, but with different rhythms than their neighbor. I would sing the root melody along with the improvising class and show with a hand sign when to change to the next pitch.

Level three focused on tonal improvisation. The students again sang the root melody on a neutral syllable, but this time instead of improvising rhythms, the students sang different chord tones on each macrobeat. For instance, if a major tonality root melody started with four macrobeats of "Do," a student might now sing, "Do, Mi, So, Mi." At the same time, another students might choose to sing, "Do, So, Mi Do." As one might imagine, this level was particularly difficult for students to master and took a lot of preparation and review of tonal patterns for each chord function.

Those students who could improvise successfully at level three were challenged to continue on to level four. In level four, students were asked to combine their rhythm and tonal improvisations. They were instructed to continue changing chord tones, but to improvise different rhythm patterns on the chord tones as they did so. Their improvisations now began to sound more like countermelodies.

The final step of the improvisation sequence was to add non-harmonic tones to the improvisation. Students were instructed to sing pitches that were outside of the given chord in between the macrobeats. Their tonal improvisations now included singing

diatonic passages instead of only singing arpeggiated patterns. At the same time, students were also challenged to continue improvising rhythmically. At this point, they were improvising true countermelodies. The five steps are summarized below:

1. Sing root melody pitches on macrobeats.
2. Improvise rhythms on root melody pitches using neutral syllables.
3. Change chord tones on each macrobeat.
4. Change chord tones while improvising rhythms.
5. Add non-harmonic tones to improvisation.

Few of the students were able to complete all five levels in one quarter, but they had a good foundation for improvisation instruction in future years. Each of the five levels can be used with a song. Once the students know a melody and the accompanying root melody, they can improvise together at any or all of the five levels simultaneously. This allows for tremendous individualization within a single class. Students are able to improvise at the level with which they are comfortable.

At the end of instruction, I administered the IPT to the students. Recordings were made of the student improvisations in accordance with normal classroom procedures. Recording performances has two distinct advantages over in-class evaluation. Since each student improvisation was to be evaluated using four different rating scales, the recordings allow me to replay student performances as many times as needed. The recording procedure also allows me to evaluate the improvisations outside of class.

The students also took HIRR and RIRR at the end of instruction to see if any progress had been made in their ability to audiate harmonic changes in music. I was curious to see how these test scores reflected a quarter of improvisation instruction. Since

this was an introductory experience in vocal improvisation for the students, neither the scores from HIRR and RIRR nor the quality of the students' harmonic improvisations were used to determine class grades. I simply focused on learning from this experience to improve improvisation instruction in the future.

Criterion Measures

Gordon (1989) designed the AMMA to provide teachers with an effective tool for determining a student's potential to achieve in music. The AMMA measures an individual's ability to audiate the tonal and rhythmic aspects of music. The test manual includes norms for junior high, high school, or college students. According to Gordon (1989), an individual's music aptitude accounts for up to 65% of the reason he or she achieves in music. Therefore, knowing a student's music aptitude enables a teacher to more effectively individualize instruction to the needs of that student. By knowing the music aptitudes of a class of students, a teacher can also plan instruction more effectively according to the strengths and weaknesses of that class.

The test takes approximately 20 minutes to administer and consists of 30 questions. Directions for the test are included on the CD and example questions are given before the main test begins. For each question, two musical phrases are played on the piano. The student indicates on an answer sheet whether the phrases are the same, different due to a tonal change, or different due to a rhythm change. Answer sheets may be hand scored or sent to the publisher for machine scoring. The three scores generated are a tonal aptitude score, a rhythm aptitude score, and a composite score. Raw scores from the test can be converted into percentiles for junior high, high school, college non-music majors and college music majors. The instruction manual includes a description of

music aptitude, a rationale for the test, directions for administering and scoring the test, suggestions for interpreting and using the results, and technical data for the test.

The test manual reports split-halves reliability coefficients of .84 for high school students. The manual also reports intercorrelations between the rhythm and tonal sub-tests from .72 to .78 and discusses implications of that data. A section covering the subjective and objective validity of the measure concludes the manual.

Gordon (1998) designed HIRR and RIRR to measure student's readiness for harmonic improvisation. An individual must be able to audiate both what the chord changes are and when they occur to successfully improvise within a harmonic progression. HIRR measures students' ability to tell whether two harmonic pattern are the same or different. RIRR measures students' ability to tell whether two time patterns (harmonic rhythms) are the same or different.

HIRR consists of 43 harmonic patterns performed in various tonalities. The patterns are all performed in the same simple rhythm. The test takes 17 minutes to administer. The recording includes directions and practice items. Students listen to pairs of harmonic patterns and mark a column for "same," a column for "not same," or if they are not sure they choose the question mark column. The harmonic patterns have been researched to determine easy, moderate, and difficult patterns to audiate. The relative difficulty of pairs of harmonic patterns has also been determined through research.

RIRR takes 20 minutes and consists of 40 pairs of recorded time patterns. The same simple melodic line accompanies each pair of time patterns. The time patterns are chord roots with a whole note, a dotted whole note, a half note, or a quarter note time value. When pairs of time patterns are different, it is because the length of one of the

chord roots has changed. Directions and practice exercises are included on the recording. After listening to two time patterns, students mark a column for "same," a column for "not same," or if they are not sure they choose the question mark column.

Raw scores for each test are converted into percentile ranks, and both scores are recorded on a student record sheet to aid the teacher in planning instruction. There are no composite scores or percentiles for the combining the scores of the two tests. The instruction manual contains a section with suggestions for how to adapt improvisation instruction to the needs of students with different levels of improvisation readiness.

The test manual reports split-halves reliability coefficients for grades 3 through 12. Reliability coefficients for high school students are reported as .87 for HIRR and .80 for RIRR. Some aspects of concurrent validity and longitudinal predictive validity for the tests have already been discussed in Chapter 2. Aspects of content and construct validity are discussed in detail in the test manual.

The other criterion measure that was administered to the students was the researcher-designed IPT. I administered this test individually. The students were already familiar with the test activity because they practiced it as a class activity before the individual tests began.

Once the student was ready to begin the harmonic improvisation test, I sang the roots to a I-IV-V^7-I progression in major tonality together with the student. The roots were sung on macrobeats with the chord changes occurring after every fourth macrobeat. Next, I asked the student to improvise a melody in major tonality and duple meter to go along with that progression. The student improvised using neutral syllables. After the student completed the activity, the process was repeated for major tonality/triple meter,

minor tonality/duple meter, and minor tonality/triple meter. In minor, the student improvised to a i-iv-V^7-i progression. I accompanied student improvisations on the piano by playing the chord roots in the bass on each macrobeat and the chords in the treble on each microbeat. An audio recording was made of student improvisations for later evaluation.

Three independent judges used these recordings to evaluate the student improvisations. The judges were music teachers with a background in teaching audiation skills, singing, and improvisation. Two were elementary music teachers in a local public school district. One was student teaching in elementary music, band, and choir at the time of this study. All three judges improvise themselves.

The judges used four five-point continuous rating-scales to measure the convergent and divergent production of each improvisation. The first two were taken from Gordon's (2001) research to determine the nature of HIRR and RIRR. These two can be considered convergent criteria because they measure whether the improvisations are musically correct as defined by the instructions for the activity. They are measuring the correct or incorrect aspects of a response.

I designed the last two rating scales for this study. These two can be considered divergent criteria because they measure the musically creative characteristics of the improvisations. They measure to what degree a student went beyond a simply correct response and improvised in a flexible, fluent, elaborate, and original manner.

The rating-scale for measuring the convergent productions of tonality, proper harmonic changes, and musical melodic contour was as follows:

1. Student sings without a sense of tonality.

2. Student sings in the appropriate tonality but without implied harmonic changes beyond tonic function.
3. Student sings in the appropriate tonality with some implied harmonic changes.
4. Student sings in the appropriate tonality with all implied harmonic changes.
5. Student sings in the appropriate tonality with all implied harmonic changes and with musical melodic contour. (Gordon, 2001, p. 22)

The rating scale for measuring the convergent productions of tempo, meter, proper time durations, and musical style was as follows:

1. Student sings without a consistent tempo and/or the appropriate meter.
2. Student sings in a consistent tempo and in the appropriate meter.
3. Student sings in the appropriate meter with some duration changes coinciding with implied harmonic changes.
4. Student sings in the appropriate meter with all duration changes coinciding with implied harmonic changes and without extending or shortening the melody.
5. Student sings in the appropriate meter with all duration changes coinciding with implied harmonic changes and with musical style. (Gordon, 2001, p. 22)

The rating-scale for measuring the divergent production factors of fluency and elaboration was as follows:

1. Student sings root melody or responds out of tonality and/or meter.
2. Student sequences a musical motive while staying within the given chord progression.

3. Student sequences more than one musical motive while staying within the given chord progression.
4. Student develops a musical motive while staying within the given chord progression.
5. Student develops more than one musical motive while staying within the given chord progression.

The rating scale for measuring the divergent production factors of flexibility and originality was as follows:

1. Student repeats root melody or responds out of tonality and/or meter.
2. Student improvises by singing chord tones other than roots or by changing rhythm patterns.
3. Student improvises by singing chord tones other than roots and by changing rhythm patterns.
4. Student improvises with chord tones other than the root in each chord function and uses at least four different rhythm functions.
5. Student improvises with non-harmonic tones while staying within the given chord progression or uses more than four rhythm functions.

Three judges listened independently to each student improvisation and assigned it a score for each of the four rating scales. Since each student performed four improvisations, each judge recorded sixteen scores for each student, four scores for each of the four criteria. I then combined these scores into a convergent production score, a divergent production score, and a total improvisation score.

Chapter Four

Results

Four kinds of data analysis helped to produce the results of this study. Test scores from this study were used to generate means, standard deviations, and reliability coefficients for the criterion measures. Intercorrelation coefficients indicated the degree of similarity between the subtests. Scores from the music aptitude and improvisation readiness tests were correlated with improvisation performance scores for evidence of predictive validity. A stepwise regression analysis gave evidence for which tests or subtests were the best predictors of successful improvisation.

Means, Standard Deviations, and Reliabilities of Criterion Measures

After scores were collected from my improvisation performance test (IPT), I correlated them according to judge to determine the interjudge reliability of the test. The judges were labeled judges A, B, and C. Upon examination of the three sets of scores, it became evident that Judge B was not clear about the definition I was using for a "musical motive" and therefore scored the improvisations quite differently from the other two judges. Judge B was dropped from the data analysis and a combined score from Judges A and C was generated. I calculated means, standard deviations, and interjudge reliability coefficients for IPT using scores from judges A and C.

The means, standard deviations, and reliability coefficients for each of the criterion measures used in this study are reported in Table 1. The first set of means, standard deviations, and reliabilities for each criterion measure was calculated using scores from this study. The second set is drawn from the AMMA test manual (Gordon, 1989, pp. 40-44) and HIRR and RIRR test manual (Gordon, 1998, p.48). Data from the

AMMA is reported for the Tonal Subtest (AMMA – R), the Rhythm Subtest, (AMMA – R), and the combined Composite (AMMA – C) score. HIRR – 1 and RIRR – 1 were the first administration of the tests, while HIRR – 2 and RIRR – 2 were the second administration of the tests. The Harmony Rating Scale (IPT – H), Rhythm Rating Scale (IPT – R), Fluency and Elaboration Rating Scale (IPT – FE), and Flexibility and Originality Rating Scale (IPT – FO) have no prior means, standard deviations, or reliabilities reported for this harmonic improvisation activity.

Table 1
Means, Standard Deviations, and Reliability Coefficients for Criterion Measures: Comparison of Present Study Results with Test Manuals

	Present Study			Test Manuals		
Criterion Measure	Mean	SD	*r*	Mean	SD	*r*
AMMA – T	24.3	4.53	-	23.8	4.37	.81
AMMA – R	26.6	4.40	-	26.8	4.03	.82
AMMA – C	50.9	8.43	-	50.6	7.91	.84
HIRR – 1	31.3	5.09	.86	28.1	5.90	.87
HIRR – 2	32.2	4.47	.65	28.1	5.90	.87
RIRR – 1	33.2	5.57	.85	27.4	5.02	.87
RIRR – 2	33.2	4.3	.79	27.4	5.02	.87
IPT – H	3.9	.97	.79			
IPT – R	4.0	.85	.67			
IPT – FE	2.7	1.11	.76			
IPT – FO	2.8	1.18	.79			

The means I calculated for the AMMA were similar to those reported by the test manual (Gordon, 1989, p.40). The split-halves reliabilities reported by the AMMA test manual were quite high (.81-.84). I was not able to generate split-halves reliabilities due to lack of access to original answer sheets at the time of data analysis.

The means I calculated for both HIRR (31.3-32.2) and RIRR (32.8-33.2) were higher than those reported in the test manual (HIRR 28.1, RIRR 27.4) (Gordon, 1998, p.51). However, when I compared means for the first administration of the tests to means from the second administration of the tests using a T-test, neither the HIRR administrations ($t=1.04$, $df=28$, $p>.05$) nor the RIRR administrations ($t=.40$, $df=28$, $p>.05$) showed significant differences. Standard deviations for both HIRR and RIRR narrowed from first administration to second administration. Once again, split-halves reliabilities as reported in the test manual were high (.87). Split-halves reliabilities generated in this study were consistent with those reported in the test manual on the first administration (HIRR .86, RIRR .85), and somewhat lower on the second administration (HIRR .65, RIRR .79). All split-halves reliabilities in this study and in the test manuals are Spearman-Brown corrected.

Interjudge reliabilities for IPT were within an acceptable range (.67-.79), but were not as high as the split-halves reliabilities of the other criterion measures. The rhythm rating scale (.67) was not as reliable as the other rating scales and may need further examination for use with this improvisation activity. Even so, interjudge reliabilities were considered to be high enough for the purpose of this study.

Intercorrelations of Subtests

One trend that appeared in the data analysis was the moderately high intercorrelations between various subtests. Gordon (1989) has reported "high intercorrelations" between the AMMA Tonal and Rhythm subtests for high school students (p. 44). This phenomenon is attributed to the nature of the test. Since there is only one test administered, the Tonal and Rhythm subtests are products of the scoring procedure. Gordon classified a question as a tonal question if two musical phrases in a question are different tonally. If they are different rhythmically, it is a rhythm question. If the two musical phrases are the same, the question is classified as both a rhythm question and a tonal question, thereby allowing twenty tonal questions and twenty rhythm questions from a thirty-question test. However, since the intercorrelations are lower than the reliabilities of either subtest, the subtests are measuring some, though not all, unique aspects of music aptitude.

Gordon (1998) reported relatively low intercorrelations between HIRR and RIRR in the test manual (p. 58). The manual also reports low intercorrelations between HIRR, RIRR, and the subtest of AMMA. However, data from this present study suggests moderately high intercorrelations among many of the music aptitude and improvisation readiness subtests. Table 2 shows the intercorrelations for AMMA, HIRR, and RIRR as generated in this study. Table 3 shows the intercorrelations for AMMA, HIRR, and RIRR as reported in the test manuals.

Table 2
Subtest Intercorrelations Generated in this Study

	AMMA – T	AMMA – R	HIRR – 1	HIRR – 2	RIRR – 1	RIRR – 2
AMMA – T	1.0	.76	.38	.25	.63	.54
AMMA – R		1.0	.52	.38	.58	.62
HIRR – 1			1.0	.36	.52	.47
HIRR – 2				1.0	.36	.52
RIRR – 1					1.0	.68
RIRR – 2						1.0

Table 3
Subtest Intercorrelations Reported in Test Manuals

	AMMA – T	AMMA – R	HIRR	RIRR
AMMA – T	1.0	.74	.28	.21
AMMA – R		1.0	.26	.24
HIRR			1.0	.32
RIRR				1.0

The intercorrelation of AMMA subtests in this study (.76) was consistent with the data reported in the test manual (.74). Intercorrelations between the AMMA – T subtest and HIRR in this study (.25-.38) were also consistent with the test manual (.28). All of the other intercorrelations in this study were higher than those reported in the test manuals.

The intercorrelation of the AMMA – T subtest and RIRR was .63 for the first administration and .54 for the second. The intercorrelation of the AMMA – R subtest and HIRR was .52 for the first administration and .38 for the second. The intercorrelation of the AMMA – R subtest and RIRR was .58 for the first administration and .62 for the second. These intercorrelation coefficients indicate that the scores from these subtests have from as little as 16% to as much as 40% in common with each other.

Table 4 shows the intercorrelation coefficients between the four rating-scales used in IPT. The Convergent Production score is a combination of scores from the Convergent Production rating-scales (IPT – H and IPT – R). The Divergent Production score is a combination of scores from the Divergent Production rating-scales (IPT – FE and IPT – FO).

Table 4
IPT Intercorrelations

	IPT – H	IPT – R	IPT – FE	IPT – FO	Convergent	Divergent
IPT – H	1.0	.93	.80	.66	.98	.75
IPT – R		1.0	.82	.74	.98	.80
IPT – FE			1.0	.88	.83	.97
IPT – FO				1.0	.71	.97
Convergent					1.0	.79
Divergent						1.0

The data in Table 4 indicates high intercorrelations among scores generated by IPT rating-scales. The correlation coefficients approaching 1.0 are expected between a score (e.g. IPT – H) and a composite score containing that score (e.g. Convergent). However, all of the individual rating-scales also intercorrelated highly with each other. The data indicate that IPT – H and IPT – FO, the rating scales with the lowest intercorrelation (.66), still have 44% in common with each other. The Convergent rating-scales had very high intercorrelations (.93), as did the Divergent rating-scales (.88).

Correlation of Improvisation Scores and Other Criterion Measures

After comparing similarities and differences of the music aptitude and improvisation readiness subtests, I correlated scores from the AMMA and the first administration of HIRR and RIRR with different sets of IPT scores. I then examined scores from the four rating-scales alone, paired them into Convergent Production scores (IPT Convergent) and Divergent Production scores (IPT Divergent), and combined them into a Total Improvisation score (IPT Total). Each of these scores were then correlated with AMMA, HIRR – 1, and RIRR – 1 scores to examine the predictive validity of the measures for successful student improvisation. Table 5 shows the results of these correlations.

Table 5
Correlations Between IPT Scores and Scores from Other Criterion Measures

IPT Scores	AMMA – T	AMMA – R	AMMA - C	HIRR – 1	RIRR – 1
IPT – H	.76***	.64***	.74***	.63***	.69***
IPT – R	.70***	.61***	.69***	.54**	.54***
IPT – FE	.72***	.63***	.71***	.38*	.65***
IPT – FO	.62***	.62***	.66***	.35*	.50**
IPT Convergent	.74***	.64***	.73***	.60***	.63***
IPT Divergent	.68***	.64***	.70***	.38*	.59***
IPT Total	.75***	.68***	.76***	.50**	.64***

*p<.05, **p<.01, ***p<.001.

All AMMA, HIRR, and RIRR scores correlated significantly with all combinations of improvisation scores. AMMA scores correlated very highly with all improvisation scores. RIRR scores also correlated highly with all improvisation scores, but not as highly as AMMA scores. HIRR scores, although correlating significantly ($p<.05$) with each improvisation score, did not correlate as highly with scores from the divergent rating-scales. In every instance, the AMMA Tonal and Composite scores correlated more highly with the improvisation scores than did either HIRR or RIRR scores. Generally, RIRR scores correlated more highly with the improvisation scores than did HIRR scores. Also, in every case except the Flexibility and Originality rating scale, scores from the AMMA Tonal Subtest correlated more highly with improvisation scores than did scores from the AMMA Rhythm Subtest.

Stepwise Regression Analysis

The next part of data analysis examined the relationships between IPT scores and scores from the other criterion measures to see which criterion measures were the best predictors of success in harmonic improvisation. I entered scores from the AMMA, HIRR, and RIRR as predictor variables in a stepwise regression analysis. AMMA Tonal, Rhythm, and Composite scores were entered as separate variables. HIRR and RIRR scores were also entered as separate variables. The last predictor variable was a composite score for improvisation readiness (HIRR + RIRR) generated by combining HIRR and RIRR scores. IPT scores were the criterion variables. Table 6 shows results from the stepwise regression analysis.

Table 6
Results of Stepwise Regression Analysis

Variables	*r*	*R*	R^2	R^2 Change
IPT – H				
HIRR + RIRR	.76	.76	.58	
AMMA – T	.60	.85	.73	.15
IPT – R				
AMMA – T	.70	.70	.48	
HIRR – 1	.42	.76	.58	.10
IPT – FE				
AMMA – T	.71	.71	.51	
IPT – FO				
AMMA – C	.65	.65	.43	
IPT Convergent				
AMMA – T	.74	.74	.55	
HIRR + RIRR	.51	.82	.68	.13
IPT Divergent				
AMMA – C	.70	.70	.49	
IPT Total				
AMMA – C	.76	.76	.57	

Surprisingly, only once was an improvisation readiness score the highest predictor of successful harmonic improvisation. The combined HIRR and RIRR score was the best predictor of success for the Harmony rating-scale. The only other variable that increased the predictive ability of the combined HIRR and RIRR score was the AMMA – T score. Together they predicted 73% of the variance in the IPT – H score.

The highest predictor for most rating-scales or groups of rating-scales was the AMMA. When a second step entered into the regression, it was usually HIRR, which has the lowest correlation with AMMA scores. Though the AMMA – C and AMMA – T scores alternated as the highest predictor variable for each IPT score, in actuality, they were almost interchangeable. This trait was due to the high correlations between AMMA – C and AMMA – T. When I removed one variable from the regression, the other entered the regression first every time as the most predictive variable.

The criterion measures were able to account for more than 40% of the variance on every rating-scale or combination of rating-scales. The criterion measures were more predictive of Convergent Production scores than Divergent Production scores. Combined AMMA – T and HIRR + RIRR scores accounted for 68% of the variance for the Convergent Production scores. Combined AMMA – T and HIRR scores accounted for 58% of the variance in scores from the Rhythm Rating scale. The AMMA – C score was the best overall predictor of success in harmonic improvisation. It accounted for 57% of the variance in the IPT Total score.

Chapter Five

Discussion

The data analysis yielded many results that were helpful in drawing conclusions regarding the problems of this study. After a brief review of the purpose, problems, and the study design, this chapter presents a discussion of the data analysis, conclusions of this study, and suggestions for music educators. Since teaching vocal improvisation may be new to some, I also report some practical insights gained during the teaching process. The chapter closes with recommendations for further research.

Research Summary

The purpose of this research was to investigate the nature of creative thinking in music by gaining information about the relationships between music aptitude, improvisation readiness and the quality of student improvisations. A better understanding of these subjects may lead to more effective improvisation instruction.

The specific problems of this study were as follows:

1. To develop a reliable criterion measure that uses both convergent and divergent production factors for evaluating student improvisations.
2. To examine relationships between music aptitude and the quality of student improvisations for evidence of predictive validity.
3. To examine relationships between improvisation readiness and the quality of student improvisations for evidence of predictive validity.
4. To examine relationships between *Harmonic Improvisation Readiness Record* (HIRR) and *Rhythm Improvisation Readiness Record* (RIRR)

scores before and after instruction for further validation of the tests as music aptitude tests.

I designed this study to address these problems using 29 members of the high school choir at Freedom Farm Christian School. Before instruction, each student completed a music aptitude test, the *Advanced Measures of Music Audiation*. They also completed an improvisation readiness test, HIRR and RIRR. Choir students then experienced systematic instruction in harmonic improvisation during warm-ups for eight weeks.

After the instruction period, each student improvised for me in major and minor tonalities and in duple and triple meters. I recorded these improvisations. They also completed HIRR and RIRR once more. Later, three independent judges evaluated the recorded improvisations using the Improvisation Performance Test (IPT) rating-scales.

The harmonic improvisation activity used in this study and the four rating-scales used to evaluate that activity constitute IPT. The Harmony rating-scale and the Rhythm rating-scale come from an earlier study and measure the convergent production aspects of harmonic improvisation. I designed the Fluency and Elaboration rating-scale and the Flexibility and Originality rating-scale for this study. They measure the divergent production aspects of harmonic improvisation. I generated means, standard deviation, and reliability coefficients for the criterion measures in this study using data from this study. Acceptable interjudge reliability for IPT was evident after correlating scores from the rating scales.

I then correlated scores from the AMMA with scores from IPT to examine for evidence of predictive validity. HIRR and RIRR scores were also correlated with IPT

scores to examine for evidence of predictive validity. Next, I correlated scores from the first administration of HIRR and RIRR with scores from the second administration of HIRR and RIRR to see if those scores would change significantly due to instruction and practice in harmonic improvisation. Finally, a stepwise regression analysis helped to clarify which criterion measures were the best predictors of successful harmonic improvisation as determined by IPT scores.

Discussion of Data Analysis

When I generated means using data from this study, one glaring difference from the data reported in the test manual for HIRR and RIRR appeared. The mean scores of the students in this study were much higher than those reported in the manual. Examining the population in this study offers one explanation for this discrepancy. The students in this study were all participants in the school music program. Many had experience with instrumental improvisation before the study began. This select, musical population would be expected to score higher on an improvisation readiness test than would the general population. It is interesting, however, that the students did not score higher on the AMMA than did the general population their age.

When I generated standard deviations using data from this study, one puzzling trait emerged. While the standard deviations for the AMMA and for the first administration of HIRR and RIRR in this study were similar to those reported in the test manuals, the standard deviations for the second administration of HIRR and RIRR were somewhat smaller. This would be expected if the means were significantly higher in the second administration, but they were not. Students who did poorly on the test the first

time tended to do better the second time, but many of the students with the highest scores the first time did not score as highly the second time.

One observation may help to explain this pattern. The students who did well the first time did not seem to be as motivated to take the test a second time. The second administration was given during exam week, and students had other pressing concerns. Nine students were seniors and were graduating that weekend. One student whose scores plummeted on the second testing had stayed up most of two consecutive nights before the test. Some of these circumstances could have also been responsible for the lower reliability coefficients generated from the second testing. It is also possible that the narrowing standard deviations could be attributed to a simple regression toward the mean on the second testing.

As stated earlier, many of the criterion measures in this study had moderate to high intercorrelations. The intercorrelations between the AMMA subtests and RIRR indicated that they have up to 40% of the variance in common. The intercorrelations between HIRR and RIRR showed that they have up to 30% in common. It is curious that these intercorrelations are so much higher than those reported in the test manuals. It may be that a more homogeneous population is once again the reason. It stands to reason that the more similar members of a population are in knowledge and intelligence, the more likely they are to score in a similar manner on a group of tests.

IPT subtests also had high intercorrelations for this improvisation activity. Perhaps the four rating-scales used in IPT should be combined into two rating-scales. It seems that little additional knowledge is gained by evaluating the performances using four rating scales. The two convergent rating-scales could be combined into one rating-

scale and the two divergent rating-scales could be combined into a second rating-scale. However, the Harmony rating-scale and the Flexibility and Originality rating-scale demonstrated many unique aspects, having only 44% in common between their scores. Another possibility would be to keep the four rating-scales separate, but to try modifying the Rhythm rating-scale and the Fluency and Elaboration rating-scale so that they measure unique aspects of harmonic improvisation.

Conclusions

The development and use of IPT in this study demonstrate that it is possible to use a criterion measure including both convergent and divergent production factors to reliably evaluate student improvisations. While the measure should continue to be refined for ease of use and higher interjudge reliabilities, it demonstrated that creative thinking in music as demonstrated by harmonic improvisation is measurable. This study also demonstrated that measuring creativity in music is possible in a high school choral setting.

Some of the most useful conclusions of this study come from the correlations between IPT scores and the other criterion measures. Successful harmonic improvisation as measured by IPT correlates significantly with both music aptitude as measured by the AMMA and improvisation readiness as measured by HIRR and RIRR. That may indicate that the musical thought processes measured by music aptitude and improvisation readiness tests are similar to the thought processes required for successful harmonic improvisation. Scores from both music aptitude and improvisation readiness tests given before instruction correlated significantly with improvisation performance scores generated after instruction. Therefore, AMMA, HIRR, and RIRR all have predictive

validity for the quality of student improvisation as measured by IPT. These tests have predictive validity for creative thinking in music.

The stepwise regression analysis sheds further light on this subject. When compared to each other, HIRR and RIRR were both effective in predicting success in harmonic improvisation, but RIRR was more successful. Surprisingly, AMMA scores predicted success in harmonic improvisation more effectively than did either HIRR or RIRR scores. The AMMA Composite score was the best overall predictor of success on IPT. A combination of AMMA and HIRR + RIRR scores most effectively predicted success on the convergent production rating-scales the best, but AMMA Composite scores alone were once again the best predictor of success on the creative thinking rating-scales.

In previous studies (Gordon, 2001) that found HIRR and RIRR the best predictors of successful improvisation, different improvisation activities and rating-scales were used. The combined HIRR and RIRR score in this study most effectively predicted the results of one of those rating scales used in previous studies, the Harmony rating-scale. The other rating-scale used previously, the Rhythm rating scale, rendered such little variation of scores in this study that it did not offer information as conclusive as the other rating-scales.

Another conclusion of this research concerns the nature of HIRR and RIRR. In this study, the tests continued to function as music aptitude tests. When considering whether a test is measuring aptitude (potential to achieve) or achievement (learned knowledge and skills), one must consider various factors. Do test scores significantly change over time due to instruction and practice? Does the test have the ability to

accurately predict success in related fields? Does the test share characteristics with other known aptitude or achievement tests? Every test exists somewhere on a continuum between a pure aptitude test and a pure achievement test. A test is predominantly a music aptitude test if it accurately predicts success in musical activities and its scores are not subject to practice and training over time. It may also show high relationships with other known music aptitude tests. Four results of the data analysis lend themselves readily to the discussion about the nature of HIRR and RIRR.

The test scores were stable despite instruction and practice in harmonic improvisation over time. Even though the students received instruction in harmonic improvisation, their scores did not change significantly between the first and second administrations of the tests. RIRR was especially stable, with almost exactly the same mean on the second administration as the first. HIRR scores were higher on the second administration, though not significantly. It would be interesting to see if the mean would have risen significantly if so many of the students scoring highly the first time had not scored lower the second time. Even though 60% of the students raised their scores on the second administration, a couple of students that dropped significantly due to external factors largely affected the mean.

Concerning the predictive ability of HIRR and RIRR, the first administration of each test before instruction correlated significantly with the quality of student improvisations after instruction. Thus, both measures had predictive validity for harmonic improvisation. HIRR was able to account for 25% of the reason students succeeded in harmonic improvisation, and RIRR was able to account for 41%. Although the AMMA

accounted for 58% and was the best predictor of harmonic improvisation success, both HIRR and RIRR did exhibit predictive validity for harmonic improvisation.

In this study, HIRR and RIRR also shared some traits in common with a known music aptitude test. HIRR correlated significantly with both AMMA subtests, although they only had from 14 to 27% in common with one another. RIRR, however, correlates more highly with both AMMA subtests. RIRR and AMMA subtests demonstrated up to 40% in common with one another. In fact, RIRR correlated more highly with the AMMA subtests in this study than it did with HIRR. That was why RIRR scores never added to the predictive ability of the AMMA in the stepwise regression analysis. The tests were too similar and the AMMA was the better predictor of the two.

This result is understandable when the types of test questions are considered. In the AMMA, students compare sets of two melodies. In RIRR, students compare sets of two time patterns (a series of chord roots). Students may tend to hear these time patterns as melodies with longer note values and process them similarly to the melodies in the AMMA. In HIRR, students compare sets of two short chord progressions. These progressions are three chords long and they start and end with a tonic chord. There is nothing melodic about the progressions and it stands to reason that students would process these very differently.

There is one discrepancy mentioned before that suggests a difference in nature between the AMMA and HIRR and RIRR. The students in this study had a mean on the AMMA consistent with a more diverse group of high school students, yet their means on HIRR and RIRR were much higher than the heterogeneous population used for the test standardization. In fact, their means for HIRR and RIRR were similar to those of college

music majors or professional music educators (Gordon, 1998). Since many of the students had experienced instrumental improvisation previously in band, one could attribute these unusually high means to the tests measuring music achievement. If this were the case, it would stand to reason that the scores would continue to rise significantly during the study. They did not.

Another possibility is that the tests are simply measuring different kinds of music aptitude. In 1958, "it was discovered that music aptitude is multidimentional" (Gordon, 1998, p. 75). The AMMA measures only certain tonal and rhythmic dimensions. HIRR and RIRR could be measuring other dimensions. Researchers have found that dimensions of music aptitude measured thus far are developmental up to age 9 (Gordon, 1989). Although Gordon reports means for HIRR and RIRR scores are stable in children as young as second grade, it could be that the dimensions measured by these tests are developmental at a younger age. The students in this study had grown up hearing part-singing in church and at school. This could have helped to develop their ability to audiate chord changes and to discern when they change. One could say that this developing ability to process chord changes is a music achievement trait, but if this were the case, means for HIRR and RIRR should have risen significantly during instruction. Therefore, the issue of higher means on HIRR and RIRR does not necessarily discredit the test as music aptitude tests.

Connections with Previous Research

The results of this study inform many questions that arose out of previous research in musical creativity. Questions regarding the nature of HIRR and RIRR have

been addressed in this chapter. Those concerning other measures of musical creativity will be considered at this time.

After examining Webster's (1987a) *Measures of Creative Thinking in Music*, I wondered if creative responses in music would correlate more strongly with music aptitude if they were limited to creative responses in tonality and meter. This study demonstrated that not only do they correlate highly, but music aptitude as measured by the AMMA was also the single greatest predictor of successful improvisation. AMMA scores accounted for 58% of the reason students succeeded in harmonic improvisation. These results coordinate with results reported by Kratus (1994) showing that audiation and certain aspects of composition are highly correlated.

After examining Gorder's (1980) *Measures of Musical Divergent Production*, I wondered if creative responses in music would correlate more strongly with music aptitude if the measure of improvisatory response were adequately reliable. Once again, this study demonstrated in the affirmative. IPT attained high enough interjudge reliability to allow for high correlations between creative musical responses and music aptitude.

After examining McPherson's (1993) *Test of Ability to Improvise* (TAI), I wondered if including descriptions along with each number in a measure of creative musical response would increase the information gained by that measure. The rating-scales used in the TAI contained one verbal description followed by the numbers one through five. While IPT rating scales contained verbal descriptions for each number, the interjudge reliabilities were not as high as those in the TAI were. The form of rating scale in IPT is more useful for communicating with and instructing students, but the

information gained from a research perspective would be greater if the reliability coefficients were higher.

Considerations for Music Education

This study affirms the connection between music aptitude and creative thinking in music. If a teacher or school district wishes to have students achieve the greatest possible success in musical creativity, they should use a music curriculum from the earliest ages that emphasizes audiation development. Since music aptitude is developmental from birth to age 9, educators should give special attention to educating parents about the importance of musical immersion at a young age. Music should be as large a part of the curriculum as possible during preschool and the early school years. Music teachers working with these students should have a background in leading students into and through musical audiation levels. Time spent on developing audiation at a young age is like investing musical currency. It is a necessary readiness for students to have success in musically creative activities throughout their lives.

This study demonstrates that successful improvisation instruction is possible in a high school performance group setting. Although vocal improvisation was a little awkward at first for many of the choir students, following the procedures outlined in this study alleviated much of their apprehension. The fact that many of the improvisation activities can occur simultaneously at levels comfortable to each individual student was both comforting and stimulating for the students. Though I did not ask students to improvise in front of the group unless they were comfortable throughout the semester, it was evident when they improvised for me that each student had attained the ability to improvise correctly and creatively at some level.

Another important aspect of this study for music educators is the fact that I was able to design and use reliable rating-scales to evaluate both the correctness of improvisations and the creativity of improvisations. Evaluating creative thinking in music does not have to be a completely subjective matter. Musical improvisations can be objectively measured and evaluated by a teacher and could constitute part of a course grade. Just as many teachers include the performance of written music as a graded assignment in class, an improvisation performance could be included as part of a more complete musical evaluation process.

The fact that there are currently standardized tests available that can predict success in musical improvisation is also useful for music educators. While AMMA was the best predictor of successful improvisation in this study as indicated by the stepwise regression analysis, HIRR and RIRR also demonstrated significant predictive ability. These tests are not intended to exclude individuals from music groups or improvisation activities. They are useful for planning the type of instruction a class will require to succeed in improvisation instruction.

For classes that score lower on these measures, more time will be needed for improvisation readiness activities such as teaching tonal and rhythm patterns in various tonalities and meters and teaching songs and root melodies by rote. Once the class has a large vocabulary of songs, root melodies, tonal patterns, and rhythm pattern in various meters and tonalities, they will have the readiness to proceed through the different improvisation levels outlined in this study. Classes with lower scores on the standardized tests will also need to spend more time learning to sing independently of others.

Ostinatos, rounds, group chord progressions and any other activity that helps students to be able to sing separate parts from those around them are valuable.

Classes with higher scores on these tests will move more quickly through the levels of improvisation. They will require less time learning to sing independently of those around them. They will benefit from moving beyond tonic, dominant, and sub-dominant harmonic functions more quickly. They will benefit from moving beyond macrobeat, microbeat, and division rhythm functions more quickly. They will be more able to attempt solo improvisations in class. They will probably enjoy improvising more than other classes and will see the improvising time as fun and motivational. Of course personality types also play into these responses, but I have observed classes with high improvisation readiness demonstrate these characteristics.

Most importantly, knowing individual test scores for students on the AMMA, HIRR, and RIRR can assist the teacher in individualizing instruction. Those that score highly on these tests may be expected to progress more quickly through the curriculum and achieve a greater degree of quality in improvisation. If a student's scores are lower on the music aptitude tests, the teacher can be ready to provide alternate instruction that will help the student achieve.

Suggestions for Teaching Harmonic Improvisation

The following are some suggestions for teaching improvisation using the method outlined in this study. Vocal improvisation leaves a student vulnerable in a personal way that other singing does not. The younger students are when they first experience improvisation, the less they tend to be intimidated. If students in high school will be improvising for the first time, the teacher should put much thought into how vocal

improvisation will be introduced. Students understand the analogy of musical improvisation being like conversing with spoken language. Teachers can explain that improvising is making up music to go along with the music of others, and that it is one of the best ways to express one's thoughts and feelings through music.

Teachers should not press for individual solo responses until the students are confident. It is better to ask for volunteers or call on students that have been improvising well during group activities. It helps greatly to have other students singing the chord progression or someone playing the chord progression on a piano while they improvise at first. When improvisation occurs in music, there is usually a group playing or singing along with the soloist. The ability to move beyond this step and improvise over an audiated chord progression without accompaniment takes a much higher level of audiation and social confidence.

Students tend to improvise rhythms on a learned root melody much more easily than they improvise chord tones other than the root. A group should practice moving around to different chord tones for each function individually before combining them. For instance, even though an individual knows that the chord tones in major tonic are DO, MI, and SO, and can sing them, it is challenging for them to sing one chord tone while others around them are singing a different chord tone. They tend to gravitate to the pitches heard around them. Having the group practice improvising tonal patterns one chord function at a time is crucial to their success. Once they are comfortable improvising tonal patterns in individual chord functions, they are ready to move on to changing chord tones on each macrobeat within a progression.

One mistake teachers may make when teaching instrumental improvisation is to let the students proceed to a higher improvisation level on their instruments before they have success with that level vocally. If students can sing it correctly, the teacher may assume they are audiating it correctly. If a students memorize that tonic function has DO, MI, and SO and figure out which keys go with those three pitches, they may or may not be audiating those pitches. Improvisation of pitches then becomes more of a technical exercise than a musically creative experience. With true improvisation, the performer is able to audiate what they are about to play or sing before they perform it. It is this kind of improvisation that holds so much value for other musical activities.

Recommendations for Further Research

This study generated many questions for further research. Some involve the improvisation readiness tests. Would another study with a similar methodology, but a different age group yield similar results? Would a similar study undertaken over a longer period further validate HIRR as a music aptitude test, or would the mean significantly rise if the length of instruction increased? Do HIRR and RIRR have significant intercorrelations with any standardized music achievement test? If so, are they higher or lower than their intercorrelations with AMMA? Would scheduling the second administration of HIRR and RIRR at a less stressful or busy time of year yield different results?

Other research questions involve IPT. Would the rating-scales in IPT achieve reliability coefficients above .80 if the procedure for explaining the measure to the judges were more thorough? Would having students improvise without accompaniment yield

different results? How would writing different Harmony and Rhythm rating-scales specifically for this activity chance the results?

Some correlations run in other musical creativity studies might yield valuable results in this study. Do years involved in music, playing a harmonic instrument such as the guitar or piano, or taking private music lessons correlate with or predict success in music improvisation? Do age, general intelligence, gender, personality type, or primary learning style significantly correlate with successful music improvisation?

As researchers and educators gain a deeper understanding into the nature of creative thinking in music, they will be able more effectively model and teach students to experience creative music making for themselves. Thinking creatively in music allows individuals to move beyond simply being a technician or artisan. It allows them to become artists.

REFERENCES

Auh, M. (1997). Prediction of musical creativity in composition among selected variables for upper elementary students. *Bulletin for the Council for Research in Music Education*, 133, 1–8.

Azzara, C. D. (1993). Audiation-based improvisation techniques and elementary instrumental students' music achievement. *Journal of Research in Music Education*, 41(4), 328-342.

Azzara, C. D., Grunow, R. F., & Gordon, E. E. (1997). *Creativity in Improvisation.* Chiccago: GIA.

Baltzer, S. W. (1990). A factor analytic study of musical creativity in children in the primary grades. *Dissertation Abstracts International*, 51(07), 2306A.

Brophy, T. S. (1999). The melodic improvisations of children ages 6 through 12; A developmental perspective [Abstract]. *Bulletin for the Council for Research in Music Education*, 142, 77-78.

Gorder, W. D. (1980). Divergent production abilities as constructs of musical creativity. *Journal of Research in Music Education*, 28, 34-42.

Gordon, E. E. (1989). *Advanced measures of music audiation* [Test Manual]. Chicago: GIA.

Gordon, E. E. (2003). *Learning sequences in music: Skill, content, and patterns.* Chicago: GIA.

Gordon, E. E. (1998). *Harmonic improvisation readiness record and rhythm improvisation readiness record* [Test Manual]. Chicago: GIA.

Gordon, E. E. (2001). *Test validity and curriculum development: Three longitudinal studies in music* (GIA monograph series). Chicago: GIA.

Guilford, J. P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Guilford, J. P. & Hoepfner, R. (1971). *The analysis of intelligence.* New York: McGraw-Hill.

Hickey, M. (2001). An Application of Amabile's Consensual Assessment Technique for Rating the Creativity of Children's Musical Compositions. *Journal of Research in Music Education*, 49(3), 235-244.

Hickey, M. & Webster, P. (1999). MIDI-based adaptation and continued validation of the Measures of Creative Thinking in Music (MCTM) [Abstract]. *Bulletin for the Council for Research in Music Education*, 142, 93-94.

Jessen, P. M. (1991). The effect of audiation on elementary students' rhythm improvisations. *Masters Abstracts International*, 31(04), 1460.

Kratus, J. (1990). Structuring the music curriculum for creative learning. *Music Educators Journal*, 76, 33-37.

Kratus, J. (1994). Relationships among children's music audiation and their compositional processes and products. *Journal of Research in Music Education*, 42(2), 115-130.

Laycock, R. P. (1992). The relationship of musical experience, musical aptitude, self-concept, age, and academic achievement to the musical problem solving abilities of high school students. *Dissertation Abstracts International*. 53(08). 2728A.

Madura, P. D. (1996). Audiation-based improvisation techniques and elementary instrumental students' music achievement. *Journal of Research in Music Education*, 44, 252-267.

McPherson, G. E. (1993). Evaluating improvisational ability of high school instrumentalists. *Bulletin for the Council for Research in Music Education*, 119, 11-20.

McPherson, G. E. (1995). The assessment of musical performance: Development and validation of five new measures. *Psychology of Music*, 23, 142-161.

National Arts Education Associations. (1994). *National standards for arts education*. Reston, VA: Music Educators National Conference.

Pogonowski, L. (2001). A personal retrospective on the MMCP: A Manhattanville Music Curriculum Project participant reflects on its innovative initiatives in light of current curriculum theory. *Music Educators Journal*. 88(1), 24-27.

Randel, D. M. (Ed.). (1986). *The new Harvard dictionary of music*. Cambridge, MA: Belknap Press of Harvard University Press.

Schmidt, C. P. & Sinor, J. (1986). An investigation of the relationships among music audiation, musical creativity, and cognitive style. *Journal of Research in Music Education*, 34(4), 160-172.

Swanner, D. L. (1985). Relationships between musical creativity and selected factors, including personality, motivation, musical aptitude, and cognitive intelligence as

measured in third grade children. *Dissertation Abstracts International.* 46(12), 3646A.

Webster, P. R. (1979). Relationship between creative behavior in music and selected variable as measured in high school students. *Journal of Research in Music Education*, 27, 227-242.

Webster, P. R. (1987a). Conceptual bases for creative thinking in music. In J. C. Perry & J. W. Perry (Eds.), *Application of research in music behavior* (pp. 158-174). Tuscaloosa: University of Alabama Press.

Webster, P. R. (1987b). Refinement of a measure of creative thinking in music. In C. Madsen & C. Prickett (Eds.), *Application of research in music behavior* (pp. 257-271). Tuscaloosa: University of Alabama Press.