BOSTON UNIVERSITY

COLLEGE OF FINE ARTS

Thesis

VALIDITY IN COMPOSITION: THE ABILITY OF FOURTH GRADE COMPOSERS TO AUDIATE COMPOSED MELODIES

by

TRAVIS MICHAEL RAMSEY

B.M., University of Southern Maine, 2003

Submitted in partial fulfillment of the

requirements for the degree of

Master of Music

2011

Approved by

First Reader

Diana R. Dansereau, PhD
Assistant Professor, Music Education

Second Reader

Jay Dorfman, PhD
Assistant Professor, Music Education

VALIDITY IN COMPOSITION: THE ABILITY OF FOURTH GRADE COMPOSERS TO AUDIATE COMPOSED MELODIES

TRAVIS MICHAEL RAMSEY

ABSTRACT

The purpose of this study was to explore the composition validity – the degree to which compositions represent the intentions of their composer – of fourth grade students. I measured the ability of participants to audiate tonal patterns that they had composed. Fourth grade students in two schools in an urban district in Massachusetts (N = 152) participated in a brief introductory lesson that included echoing, improvising, composing, and discriminating between tonal patterns. Then, all participants completed the Intermediate Measures of Music Audiation (IMMA) (Gordon, 1986). Finally, a stratified random sample at each school was selected to complete the researcher-designed Discrimination of Sameness or Difference test (DSD).

Two results were significant. First, participants were able to correctly determine sameness or difference between a self-composed melody and a researcher's performance of that melody at an accuracy level of 83% on the DSD measure, and this score was significantly different than a chance score ($p < .001$). Based on this, I made the implication that audiation-based composition is the best pedagogy for composition. Second, there was a moderately weak ($r = .258$) yet statistically significant ($p < .05$) correlation between participant tonal IMMA scores and DSD scores. Based on this, I made the implication that, while music aptitude does affect participant DSD scores, students should never be excluded from composition activities based on a low music

aptitude score. Other results approached significance. First, participants from School A outperformed students from School B ($p = .07$). Second, participants seemed more likely to score correctly when response types were *same* or when the intervallic content of the composed melody contained at least one melodic third, but the difference was not statistically significant. Based on this, I recommended that composition activities begin with melodies containing large contours and introduce smaller contours as students develop. I suggested further research that would help determine if a more difficult DSD measure would produce more significant results. Also, I suggested several possible follow-up studies that would allow researchers to measure to what degree students can audiate more complex patterns.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Music composition is an essential component of any effective general music curriculum (MENC, 1994; Massachusetts Department of Elementary & Secondary Education, 1999). Beyond the simple objective of teaching students how to compose, the musical lives of students are often enriched in other areas when they study composition. For example, Azzara (1993) and Stoltzfus (2005) found that experiences composing improved students' overall music achievement. The profound effect of teaching children to compose music can even extend beyond their musical lives. For example, composition instruction can help to foster a more positive learning environment (Strand, 2005), provide a much-needed expressive outlet for at-risk teenagers (Hickey, 2009), and deeply impact students with severe disabilities (McCord, 2009). For these reasons and many others, a substantial quantity of music educators believe that teaching composition is important and valuable. For example, nearly 90% of music educators in one survey indicated that they teach composition to some degree (Strand, 2006).

Rationale for Study

Although there are many different ways to teach composition, I will discuss two in detail in the present study. First, many educators teach composition by focusing primarily on the writing of notation (see Dammers, 2007; Prasso, 1997; Strand, 2005). Composition activities in these notation-based pedagogies often include tasks like composing melodies in standard notation, and the focus is on creating a musical product in a concrete form (usually on paper). Second, other educators (see Azzara, 1993; Hagen, 1996; Stoltzfus, 2005) base their composition pedagogy on audiation. Activities in audiation-based pedagogies include building students' tonal and rhythmic vocabulary through echoing and improvising. The focus of audiation-based composition is on the process of creating music, and the final completion of a composition in its concrete form is a secondary objective. According to Han (2004), university-level composition teachers often disagree about whether composition should be taught in a notation-based or an audiation based setting.

This larger question of which approach is better for composition is beyond the scope of the present study. Instead, I will focus the discussion on informal observations of my own students. For example, I have found anecdotally that when I have asked students to compose simple melodies by writing notes on the treble clef – notation-based composition – they are able to do so with some success by the fourth grade. When asked to sing, play or identify those melodies, however, they are often unable to make the mental connection between sound and sight. This led me to believe that there is a disconnect between what young composers write on the page and what they hear in their imaginations. The students seemed proud of their compositions, but they would have been unlikely to recognize their own music if it were performed for them to hear. I developed the present study to explore this problem further, to find out whether my students were composing what they thought they were composing.

Purpose of the Study

The purpose of this study was to explore whether fourth graders are able to audiate music that they have composed. An additional purpose was to identify any relationships between that ability and music aptitude, age, sex, and school assignment.

Research Questions

In order to address the purpose of this study, I investigated four research questions. *(1) Are participants able to audiate tonal patterns that they have composed?* I employed a researcher-designed Discrimination of Sameness or Difference (DSD) measure to determine whether participants were able to audiate their composed tonal patterns. Then, I compared the participants' scores to a chance score of 50% correct responses. The null hypothesis was that participants' responses on the DSD measure are not significantly different than chance responses. The alternative hypothesis was that participants' responses on the DSD measure are significantly different than chance responses.

(2) Is there a relationship between participants' abilities to audiate given tonal patterns and their ability to audiate patterns that they have composed?

I analyzed participants' scores on the Intermediate Measures of Music Audiation (IMMA) test (Gordon, 1986) and participant scores on the DSD test to determine if there was a significant relationship between participants' music aptitude and their ability to audiate their own composed patterns. The null hypothesis was that there is no significant correlation between participants' DSD and IMMA scores. The alternative hypothesis was that there is a significant correlation between participants' DSD and IMMA scores.

(3) What effect do age, sex, and school assignment have on the ability to audiate self-composed tonal patterns?

I analyzed the effect of participants' age, sex, and school assignment on their DSD score. The null hypothesis was that none of these variables significantly affect participants' DSD scores. The alternative hypothesis was that the participants' age, sex, and/or school assignment significantly affects their DSD scores.

(4) What effect do the intervallic content (such as the presence or absence of unisons, seconds, and thirds) and response type have on the ability to audiate self-composed tonal patterns?

I categorized a sample of participants' composed tonal patterns into six mutually-exclusive categories depending upon the patterns' intervallic content (IC). In addition, I analyzed the number of correct responses for each response type – whether the two prompts played for participants to compare were the same or different. I then analyzed these groups of patterns to determine if participants were equally likely to correctly answer all responses regardless of interval content or response type. The null hypothesis was that there is no significant difference in the number of correct responses for melodies with different intervallic content or prompts with different response types for the participants who completed the study. The alternative hypothesis was that there is a significant difference in the number of correct responses for melodies with different intervallic content and/or prompts with different response types for the participants who completed the study.

Background

An understanding of audiation was fundamental to the present study. Audiation is the ability to "hear and comprehend music for which the sound is no longer or may never have been present" (Dalby & Gordon, 2009, p. 5). According to Dalby and Gordon, audiation is learned in two stages and several underlying phases. The beginning stage of learning audiation is called *discrimination learning.* In discrimination learning, students echo patterns from their teacher in order to build a tonal and rhythmic vocabulary. The first phase in discrimination learning is the creation of an "aural/oral loop" (p. 9) where students learn to echo simple tonal and rhythmic patterns. When students are capable of assigning names (such as solfege or rhythm syllables) to these patterns, they are performing at the second phase, *verbal association.* Students who are more experienced will be able to perform at the third phase, *partial synthesis:* giving tonal or metric context to patterns they hear. Only once these first benchmarks are met can students be expected to attempt the fourth phase, *symbolic association:* bringing the meaning they have learned in their experiences to the notation they are presented. *Composite synthesis* is the fifth and final phase: being able to read and write tonal and rhythmic patterns and identify their meter and tonality while audiating.

The more advanced stage of learning, according to Dalby and Gordon (2009) is *inference learning*. Instead of learning primarily through echoing patterns, students at this level expand their vocabulary by exploring unfamiliar patterns without direct teacher instruction: "Whereas in discrimination learning a teacher teaches a student both *what* to learn and *how* to learn it, in inference a teacher teaches a student only *how* to learn. The student teaches himself *what* he learns" (p. 10). The first phase in inference learning is *generalization*, and it has three sub-levels: the ability to discriminate between familiar, previously-learned patterns and unfamiliar patterns (generalization-aural/oral), apply solfege syllables to unfamiliar patterns (generalization-verbal), and the ability to sing unfamiliar patterns by sight when presented in notation (generalization-symbolic). The second phase is *creativity/improvisation*. In creating music, students must be able to fit their learned tonal patterns into a given tonal or rhythmic context. The third and final phase is theoretical understanding, where students are expected to learn standard notation. Dalby and Gordon stressed the importance of saving music theory for the end of the sequence when all the previous phases have been mastered, and warned that introducing theory too soon "can only hinder audiational development" (p.11).

It is noted that in all of their work, students should be bringing meaning to the notation, not taking meaning from it. If they are audiating correctly, the notation will simply help them to recall a pattern that they already know. Likewise, audiation-based composition calls upon students to apply the patterns they know to their creative product, much like young writers apply the words they know to a new story they are writing. Audiation ability is synonymous with music aptitude, and is commonly measured by several tests including the Intermediate Measures of Music Audiation (IMMA) (Gordon, 1986).

For the purposes of this study, I defined composition validity as the extent to which composers are composing what they are audiating. For example, compositions written by those who are able to fully comprehend the resulting sound of each pitch, and faithfully reproduce those pitches in concrete form (either by repeated performance or in some sort of notation) are examples of high composition validity[1]. These composers would be likely to recognize their compositions in performance and/or be able to identify pitch errors in those compositions.

[1] The use of this term was suggested by J. Dorfman to describe this phenomenon.

Chapter 2

Review of Related Literature

The way that children compose music is the subject of a substantial body of research. This research is most easily understood when divided into three categories. First, researchers have elucidated the creative process underlying composition by defining, describing, and quantifying the creative impulse. Second, researchers have described how the ability to audiate tonal and rhythmic patterns allows young composers to contextualize musical elements and build vocabulary. Third, researchers have examined in great detail the act of composing and the resulting musical products. These three categories are reviewed in this order, beginning with the genesis of creative ideas, proceeding through the understanding of audiation and the development of tonal and rhythmic vocabulary, and culminating with an analysis of research about young composers.

The Genesis of Creative Ideas

Kratus (1989) investigated how students managed their time when engaged in a composition task. After a brief warm-up introducing steps, skips, and leaps, 7-, 9-, and 11-year-old children (N = 60, n = 20 for each age group) were asked to create a "brand-new song, one that no one has ever heard before" (p. 9). Additionally, participants were expected to complete a repetition task – to be able to play their song twice on an electronic keyboard – which introduced a component of memory into the study. Participants were given 10 minutes to compose and practice their song. Kratus found significant differences ($p < .01$) in the strategies employed by different age groups. The

7-year-olds spent the most time exploring the keyboard but very little time developing or repeating material, while the older groups explored less and developed material more frequently. The 11-year-olds used significantly ($p < .001$) more repetition in their compositions than the 7-year-olds. Kratus claimed that the younger students directed their attention towards the process of composing, but the creation of a final musical product was a secondary concern for the younger students. As a result, they were unable to complete the repetition task successfully. In contrast, the older students were able to complete the repetition task because they allotted more time in their composition processes to practice for it. Kratus gave several possible explanations for the younger students' inability to complete the repetition task. First, they may have been unable to simultaneously hold a melody in their memories and develop it. Second, they may not have learned the proper strategies for developing their compositions or they did not understand what processes would be necessary to complete the task. Third, Kratus underscored the importance of the younger students' focus on process instead of product. In summary, Kratus found that the youngest composers approach composing largely through exploration. As they grow and develop, their compositions – and the way they comprehend them – begin to fit into the tonal, rhythmic, and formal constructs that resemble those of adult composers.

In a 1995 mixed-method dissertation, Hickey sought to answer those questions about process and product. Hickey examined the relationship between the musical thought processes and products of fourth and fifth graders (n = 21) engaged in a composition task. To gain an understanding of participants' creativity, Hickey collected quantitative data representing measures of the students' aptitude and achievement and specific details about their compositions. Hickey measured their creative music aptitude (measured using Webster's MCTM -II (1987, 1994)), level of performance experience, and their teacher's ratings of their creativity using a researcher-designed survey instrument. Then, the participants created compositions using a computer program with unobtrusive MIDI recording capabilities. Hickey analyzed the MIDI files for fluency: "time spent in each section, average length (measure) of MIDI files", "average number of notes (keystrokes) played," and "average measures of music versus measures of silence" (p. 107). Hickey then analyzed the files for evidence of flexibility: "average range of notes," "average number of timbres used", and "average number of changes in the parameters of high/low, soft/loud, and fast/slow" (p. 107). Hickey determined the level of craftsmanship and creativity using Amabile's consensual assessment technique (1982). Hickey found that interjudge reliability for the evaluation of craftsmanship was higher than for that of creativity.

Hickey (1995) found several significant results among the quantitative data. Hickey grouped the participants' compositions based on creativity and craftsmanship, rating each composition as high, average, or low. Many of the high creativity compositions were also rated high craftsmanship. First, students whose compositions were placed in the high creativity group had significantly more parameter changes and a higher ratio of playing time versus silence than their peers in the lower group, although the significance level was rather large ($p < .10$). None of the other variables coded in analysis of the MIDI files varied significantly between high and low groups, nor was there a significant difference between flexibility and fluency among groups. Second, there were also differences among the craftsmanship groups. Hickey found that students in the high craftsmanship group composed fewer measures, but with more keystrokes overall than the low craftsmanship group. Like the creativity groups, there were no significant differences between fluency and flexibility among the craftsmanship groups. Third, Hickey found no significant correlation with process variables (time, measures, etc.) and participants' scores on the MCTM-II (Webster, 1987, 1994) test, a finding contrary to what had been expected. Fourth, Hickey found that only a few process variables correlated significantly with the teacher's ratings of student creativity. Hickey suggested that this was due to teachers being unlikely to be able to recognize their students' creative potential.

Hickey (1995) also employed qualitative techniques in describing the participants' processes and products. First, Hickey identified several general descriptors of participants' approaches: "literal", "classical," "rebel," "non-creative," "fluent motivic," "development motivic," and "physical" (p. 146). Second, Hickey discovered that participants employed different levels of development, exploration, and repetition in their compositions. Third, Hickey described *composition emergence*, which revealed whether aspects of the final composition could be identified in the "introduction" session, the "exploration" session, or the final "composition" (p. 148) section. Fourth, Hickey found that several participants seemed to return to particular "recurring musical patterns" (p. 149). The qualitative findings of seven participants whose compositions were rated as low creativity were compared with those of six participants whose compositions were rated as high creativity.

Hickey (1995) found several important differences between composers whose work was analyzed qualitatively. Composers of low creativity compositions were not observed to develop or experiment with motives. Even composers with piano training, who had a repertoire of skills to draw from, were unable to develop musical ideas. There was little use of repetition in the low creativity group as well. Composers in the low creativity group, when challenged to employ dynamic changes, often did so using a single note. Another difference between the creativity groups was that high creativity compositions emerged comparatively late in the process. Low creativity compositions were less consistent, but often emerged very early or not at all.

In summary, Hickey (1995) demonstrated that there are substantial and important differences in the ways more or less creative students compose music. In general, Hickey claimed that high creativity students demonstrated both more fluency and flexibility and more development and exploration than low creativity students. This corroborates the findings of Webster (1979), who found that fluency and flexibility led to more divergent thinking skills. The lack of quantitative findings regarding creative aptitude, teacher ratings, and indicators of fluency and flexibility in Hickey's study highlights the difficulty of assessing students' creative potential, both with informal surveys and published tests. The relatively low interjudge reliability for creativity scores also demonstrates this.

Webster (1977, 1979) is responsible for much of the foundational research on musical creativity. The Measures of Creative Thinking in Music test (MCTM), an instrument that measures the creative potential of students by quantifying their musical extensiveness, flexibility, originality, and syntax, was developed by Webster (1987). In a validation study of the MCTM, Webster further concluded that creativity and audiation are separate and independent abilities.

In a study related to those of Kratus (1989) and Hickey (1995), Auh (1995) sought the best predictors of creativity during a composition task. Auh investigated the effect of music experiences on fifth and sixth grade students (n = 67) and found that 25% of the variance in compositional creativity could be explained by students' informal music experience, music achievement, and academic grades. Based on the finding that informal music experience accounted for 11% of the total variance, Auh came to the conclusion that students' involvement in music outside of school (e.g. their informal music experiences) is crucial to developing their creativity.

In conclusion, the creative process of young composers begins with divergent thinking and exploration, continues through repetition and development, and concludes with the convergence of ideas into a final product. Even though all of the individuals in the reviewed studies responded differently to their tasks, there are substantial differences between groups of subjects. Webster (1979) found that students who were able to work with fluency and flexibility were more adept at divergent thinking. Hickey (1995) found that more creative students tended to develop their ideas more fully, causing their finished compositions to emerge later in the process than their less creative peers. Kratus (1989) found that younger composers were, in general, likely to spend more time exploring and less time developing. This verifies Hickey's finding that less creative students demonstrate less development in their compositions, especially considering the fact that creativity develops with age. These themes will continue to be important in the next section, where students' ability to audiate tonal patterns will be examined.

The present study is similar to the research discussed in this section in several important ways. Participants were given time to explore and improvise before the composition task begins. Also, as will be seen in the explanation of the method in the next chapter, the participants' composition task was primarily process-centered. The participants did not create a concrete musical product, rather they participated in a guided activity where they explored musical possibilities. An analysis of the effect of participants' musical activities outside school is beyond the scope of the present study.

Audiation and the Learning of Tonal Patterns

Kratus (1994) investigated the relationship between audiation and many of the process and product ideas that appeared in Hickey's (1995) dissertation. Using the same instructions as a previous study (Kratus, 1989), Kratus instructed 9-year-olds (n = 40) to spend 10 minutes composing a new song. Participants used a synthesizer connected to a computer with MIDI recording capability similar to that used by Hickey.

Kratus (1994) analyzed the composed works quantitatively to learn about the participants' processes and products. To analyze process, two judges listened to the entire 10-minute session and coded each 5-second interval as exploration, development, repetition, or silence. Two other judges rated the musical products composed by the participants using a researcher-designed instrument. Judges rated each composition on the basis of its tonal cohesiveness, metric cohesiveness, and the absence or presence of tonal and/or rhythmic patterns (and whether the patterns were merely presented, repeated, or developed). Interjudge reliability among product and process ratings ranged from .83 to .95. Additionally, Kratus determined the extensiveness of the compositions by recording the range of pitches used and the length of the song in seconds.

Kratus (1994) found several significant results among process, product, and audiation variables. First, Kratus found significant correlations between audiation skill and the creative process used. For example, composite aptitude scores were negatively correlated with exploration ($r = -.46$, $p < .01$). Tonal aptitude was positively correlated with the use of development ($r = .34$, $p < .05$). Rhythmic and composite aptitude were positively correlated with the use of silence ($r = .32$ and .33, respectively, $p < .05$). In summary, participants with higher aptitudes generally spent less time exploring, more time developing, and more time silently working out compositional problems. According to Kratus, these correlations indicate that audiation enabled students to think about their ideas without needing to find them through exploration. Second, Kratus found correlations between audiation and the final products composed. Composite aptitude scores were positively correlated with tonal and metric cohesiveness ($r = .39$, $p < .05$; $r = .45$, $p < .01$, respectively). Rhythmic and composite aptitude scores were negatively correlated with pitch range ($r = -.33$ and $-.36$, respectively, $p < .05$). Participants with higher aptitudes generally created more cohesive compositions that used a narrower pitch range. Kratus suggested that the likelihood of lower aptitude participants to use a broader pitch range was evidence that their compositions were less fettered by traditional notions of tone and contour. Third, Kratus found significant correlations between the processes used and the products created. For example, development (r = .47, p <.01), silence (r = .36, p < .05), and repetition (r = .51, p < .001) were positively correlated with tonal cohesiveness, while exploration was negatively correlated with tonal cohesiveness (r = .69, p < .001). Participants who spent more time developing, silently thinking, and repeating materials created more cohesive compositions, while participants who spent comparatively more time exploring created less cohesive compositions. These findings are parallel to those of Kratus (1989), but in this (1994) study the participants' compositions more closely resemble adult compositions as music aptitude, not age, increases.

In summary, Kratus (1994) found important evidence that audiation skill plays a critical role in the creation of new, cohesive musical products. Although the correlations among audiation, process, and product were moderate, the study provided a foundation for later researchers who would explore the ways the skill of audiation provides composers with context and vocabulary.

Hagen (1996) verified many of the findings of Kratus (1989). Hagen sought to examine the role of audiation on the creative processes and products of fifth-grade students (n = 38). Like the Kratus study, participants were given time to explore the sounds before creating their compositions – all of which was recorded through MIDI software. The task culminated with participants performing their finished compositions from memory. The creative process of the participants was analyzed for four variables (exploration, repetition, development, and silence), and the products composed were analyzed for cohesiveness and the use of repeated and/or developed rhythmic patterns.

The study differed from Kratus's study in three ways. First, Hagen's participants composed using an electronic keyboard set up to play 10 [unpitched] percussion timbres on 10 adjacent white keys instead of an electronic piano sound. Second, Hagen interviewed participants after the activity to discover whether their compositions "sound[ed] the way they were intended" (p. 5). In the context of the present study, this is an example of gauging participants' composition validity. Third, Hagen gave participants only 5 minutes to compose instead of the 10 minutes given by Kratus. Hagen reported that the judges had difficulty telling the difference between development and exploration; interjudge reliability for these variables was only r =.57. It seems possible that the timbral difference and the lack of tonal context made this task more difficult for the judges.

Many of Hagen's (1996) findings were similar to Kratus' (1989, 1994). Like Kratus, significant correlations were weak to moderate. Hagen found negative correlations between rhythmic audiation and amount of exploration (r = -.33) and composite audiation and amount of exploration (r = -.35)[2]. Hagen suggested that this relationship is evidence that greater audiation ability leads to a lessened need for exploration. In contrast to Kratus (1994), no other significant correlations were found. Another contrast to Kratus' findings is that 66% of the participants' time was spent in repetition or development (in Kratus's study, 63% of the time was spent exploring). It seems likely that the lack of correlation and difference in time spent could be a result of older students or tighter time constrictions in Hagen's study. What is important to note is

[2] A *p* value was not reported by Hagen.

that, even considering the differences in the two studies, Hagen confirmed the inverse relationship between audiation ability and time spent exploring.

During the interview following the activity, 58% of participants reported that their compositions sounded the way they intended. There are two plausible explanations for the moderate proportion of students who reported high composition validity in their compositions. First, it seems likely that students find it easier to tell the difference between different timbres than different pitches. For example, a student playing the pitches *c d e* on a keyboard set up to sound like a piano as in Kratus' (1989) study would hear *do re mi*, which could easily be confused with *fa sol la*. A student who plays the same keys in Hagen's study, however, would hear three distinct percussion instruments. A second explanation is that development is far more difficult when the composer is unable to use pitch. Devices like sequences, which sound logical on a pitched keyboard fail to make musical sense when each key is a new, unpitched, timbre. Participants who struggled with composition validity cited failings of their own memories, errors in their performance, and a lack of exploration time as explanations. It seems likely that students were over-stimulated by the wide selection of timbres, especially if they had difficulty developing their compositions in an unpitched environment. The finding that participants reported needing more exploration time suggests that Kratus' 10 minute session is more appropriate for a composition activity involving memorizing a performance. In the present study, I avoided many of the difficulties encountered by Hagen by performing all composition activities in an identical timbre. Also, the allotted time for each participant's individual composition session in the present study is 10 minutes – the same as that used by Kratus.

Josuweit (1991) also examined the role of audiation, exploring whether a researcher-designed audiation-based music curriculum would affect the creativity of fourth grade beginning band students (n = 21). Josuweit used a quasi-experimental format, dividing the participants into two groups. Group assignment was based on scheduling necessities. One group received traditional instruction; the other learned in an audiation-based approach. After 20 weeks of instruction, the students' music aptitude and creativity were measured.

Josuweit (1991) did not find any significant differences between the control group and the experimental group. It is important to note, however, that Josuweit did find a significant negative correlation between rhythm aptitude and music creativity. In discussing these findings, Josuweit claimed that the available creativity tests do not allow researchers to discriminate between audiation ability and engaged exploration. Josuweit then called for audiation-based creativity tests.

In summary, according to researchers, the ability of students to audiate (their music aptitude) is essential to their development as young composers. Kratus (1994) found that students with high music aptitude were able to spend less time exploring and more time developing their musical ideas, a finding that Hagen (1996) confirmed. The importance of audiation in creativity was also summarized by Dalby & Gordon: "In order to create or improvise, the student must have something to create or improvise with. The tonal and rhythmic patterns learned … comprise the content the student uses to form his own unique musical ideas in creativity and improvisation" (2009, p. 11).

The findings described in this section lend credibility to my analysis of correlation between the participants' music aptitude and their scores on the audiation task in the present study. For example, Kratus (1994) found that participants with higher music aptitude were better able to repeat and develop motives in their compositions. Higher aptitude participants were also able to contemplate their motives in silence. Participants in the present study with higher aptitude who are able to hold motives in their memories while comparing them to performed prompts seem likely to perform better, which will result in a positive correlation between music aptitude and audiation task scores.

Music Composition in General Music

Despite the body of research detailing what is understood about creativity and how music aptitude affects creative processes and products, creative music composition activities are hardly a feature of general music curricula in the United States. A 2006 survey showed that there is little agreement among K-12 music educators as to the definition, value, and appropriate pedagogy of composition (Strand). Nearly 90% of respondents (N = 349)[3] claimed to teach composition to some degree, but less than 6% claimed to teach composition "often" (p. 159). This seemingly low percentage could be due to bias on the survey instrument: Strand coded the five Likert-type responses as "often, sometimes, rarely, very rarely, and never," (p. 159). The placement of "rarely" in the middle position of the ranked responses may have led participants to believe that a "sometimes" response was unusual. According to Strand, teachers differed in how they

[3] This figure differs from Strand's reported figures (137 ensemble teachers and 212 general music teachers for a total of 349 respondents).

valued the teaching of composition. More than two-thirds (71%) of the respondents who teach composition felt that “children learn more through composing,” while more than half (65%) valued composition as an opportunity to “enrich other learning” (p. 158). Strand also reported that teachers had many different opinions about the definition of composition and the goals they hoped their students would achieve through its study. Goals reported included developing skills in improvisation, self-expression, critical thinking, technology, and learning to appreciate famous composers. Only two respondents (representing less than 1% of the sample) taught composition with the specific goal of developing the skill of composing. This finding is at least somewhat confirmed in the research (for example, Stoltzfus’s (2005) use of composition to increase performance achievement). Among teachers who did not teach composition, the most common reasons were having too many other important things to teach and not having access to technology.

Along with differing in how composition was defined and valued, respondents reported using a variety of pedagogical methods. Strand defined three types of composing tasks: heavily structured (students were told which pitches and rhythms they could choose), moderately structured (less tonally-directed compositions where students were given broad guidelines such as ‘ABA form’), and unstructured tasks (students were encouraged to compose, but given no specific directions or assignments). No explanation was given for the finding that none of the reported assignments included revision or reflection. Perhaps the lack of reflection is due to the fact that most respondents were using composition to teach non-composition concepts. It is also possible that Strand had not directly asked the respondents to comment on the role of reflection. Strand closed by suggesting further research on composition pedagogy and examining what and how students learn under different levels of structure.

Further demonstrating Strand's (2006) claim that educators value composition differently, Stoltzfus (2005) saw composition as a possible method for increasing musical achievement of band students. Stoltzfus investigated whether instruction in audiation-based composition would increase the instrumental music achievement of fourth graders (n = 64). In a study similar in scope and design to that of Josuweit (1991), half of the participants received audiation-based instruction that included composition activities. A control group also received audiation-based instruction, but without experiencing composing. Both groups completed Gordon's Music Aptitude Profile (1995) before the instruction period and a researcher-designed achievement measure after the instruction period.

Stoltzfus (2005) reported that interjudge reliabilities for the researcher-designed measure ranged from .645 to .980, with the exception of two outliers (.291 and .450). The etudes performed by participants who were assessed by the measure proved too difficult, as Stoltzfus reported that scores were significantly lower than the expected mean. Still, Stoltzfus reported that the experimental group scored significantly higher than the control group on both the researcher-designed achievement measure in overall singing achievement ($p < .005$) and a composite score for overall singing and playing achievement ($p < .05$). Stoltzfus (2005) also reported qualitative findings based on a 5-question interview given after the final measure. Four students from the experimental group were chosen for interviews based on “their potential to provide rich, detailed information and their perceived comfort with the interview process” (p. 77). Participants reported using audiation in various ways when performing, including using pattern recognition to audiate notated pitches before playing them. It was not reported how the interviewed students scored on the aptitude or achievement tests, but the clarity of the responses and their seemingly firm grasp of audiation suggests that the interviewed students represented the stronger musicians.

Guderian (2008) helped to quantify the role of music composition in general music classes by measuring the effect of composition and improvisation activities on participant achievement in sight-reading, recorder performance, and music theory. Fifth grade participants (N = 46) in two heterogeneously grouped classes took pretests on sight-reading, recorder performance, and music theory. Participants also completed the IMMA (Gordon, 1986). IMMA scores were used to ensure that the two groups were equivalent in music aptitude. Over the 18-week study, Guderian instructed each of the two classes using the same curriculum content and method of delivery. The two groups differed in how students reviewed and reinforced their learning: a control group practiced through the use of “drill and practice techniques common to music education,” (p. 170) while the experimental group performed six different composing and improvising tasks. At the conclusion of the study, participants in both groups completed posttests.

Guderian (2008) reported that nearly all participants’ scores increased dramatically (no significance level was reported) from pretest to posttest, demonstrating that the instruction was effective. There was not a significant difference in achievement between the two groups on the pretest or posttest scores at the .05 level when pretests and posttests were compared using two-tailed t-tests. Based on these findings, Guderian concluded that replacing standard classroom practice activities with composing and improvising activities did not negatively affect student achievement. Guderian also reported that students in the experimental group genuinely enjoyed their composition activities. It should be noted that Guderian's study did not include instruction in audiation; the composition activities were notation-based rather than audiation-based. Guderian's composition activities were similar to many of those described by Strand (2006). Little emphasis was placed on revision, and the ultimate goal in including composition in music education was an increase in another area of music achievement and not increasing composition skill.

Dammers (2007) echoed Strand's (2006) claim that many educators find it difficult to fit composition instruction into an already-busy curriculum. Dammers instructed eighth grade students (n = 24) in a middle-school band program to spend 14 weeks composing with the aid of laptops equipped with notation software. Unfortunately, a separate, quiet room was not available and the young composers were forced to compose wearing headphones in the back of the rehearsal space during rehearsal. Dammers was seeking evidence that students could transfer their understanding of specific aspects of a model composition into their new creations. Like other studies that came before (Hickey, 1995, for one), Dammers evaluated their final products for craftsmanship and creativity. Dammers also rated the compositions for evidence of the composers' transfer of characteristics of the model compositions.

Dammers' (2007) interjudge reliability was moderate (ranging from .624 to .830). The compositions' craftsmanship and creativity was characterized by Dammers as "fairly modest" (p. 73), with mean scores of less than four on a Likert-type scale that ranged from zero to eight. Participants were generally unable to transfer their understanding of characteristics of the model composition; the mean score for conceptual understanding was 2.66 (also on a Likert-type scale ranging from zero to eight). A closer examination of the compositions revealed that many reflected an "exploration of the Finale NotePad interface, rather than musical intent" (p. 89). This issue – a threat to composition validity – was also noted by composers interviewed by Han (2004) below. Based on the moderate levels of craftsmanship and creativity, the substantial difficulty of listening to a laptop through headphones during a band rehearsal, and the added challenge of learning a new software interface while first exploring composition, participants in Dammers's study were generally unable to compose creatively.

In a related study, Prasso (1997) investigated the differences between computer-based composition and traditional pencil-and-paper composition. In contrast to Dammers' (2007) study, Prasso relied on older technology and conducted the study without the disruption of a concert band rehearsal. High school students enrolled in a choral program ($n = 60$) spent a 45 day period composing, 30 in a control group with paper and 30 in an experimental group using keyboards and computers. The sight-singing ability of both groups was measured with an identical pre- and posttest.

Interjudge reliability for the researcher-designed sight-singing measurement was very high ($r = .99$). Prasso (1997) found that there was a significant increase ($p < .05$) in the sight-singing ability of students in the experimental group, but no such increase in the control group. Prasso claimed that a major reason for the findings may have been the participants' access to immediate visual and aural feedback of their composition by the computer – an opportunity not present for traditional composers. The value of such feedback was also noticed by several of the composers interviewed by Han (2004). Prasso also attributed the difference to a higher level of motivation among participants in the experimental group. This motivation was primarily the result of the immediate auditory and visual feedback that is common in computer learning environments.

Prasso (1997) and Dammers (2007) investigated the value of composition for teaching other musical skills (sight-singing ability and music performance achievement, respectively). This line of inquiry is related to Strand's (2006) finding that music composition is seldom taught with the purpose of teaching composition skills. To investigate the actual development of composing skills, Wilson and Wales (1995) examined the compositions of second grade (n = 40) and fourth grade (n = 40) composers to determine if certain characteristics could be learned about each age group. Similar to studies by Kratus (1989, 1994), students were given 10 minutes to compose. Participants used a computer program in an isolated room. Unlike other related studies, however, (Hagen, 1996, Hickey, 1995), no electronic keyboard was used. The researchers observed the students for four qualitative variables: the method of composition, the use of the playback mechanism, the use of the eraser, and their level of familiarity with the computer and mouse. Compositions were evaluated and rated for their melodic and rhythmic complexity in three stages. Stage one melodic complexity indicated notes that were placed primarily for their appearance on screen, with no evidence of the use of melodic contour [likely no composition validity because students would have no aural memory of their composition], stage two melodic complexity indicated a melodic contour composed of steps and small leaps [likely little composition validity], and stage three melodic complexity indicated a composition with both melodic contour and organization around a single tone [likely higher composition validity due to the concentration around a central tone]. Similarly, stage one rhythmic complexity was based on visual appearance with no apparent grouping of rhythms, stage two complexity indicated a composition that included rhythmic phrases and deliberate combinations of rhythmic values, and stage three complexity indicated a composition with a clear sense of meter underlying rhythmic groupings.

Reliability for the two judges ranged from .69 to .92 for melodic complexity and .49 to .94 for rhythmic complexity. Due to this relatively low interjudge reliability, Wilson and Wales (1995) did not treat any of their findings to be statistically significant. They suggested that further research would need a clearer distinction between stage one and stage two to produce reliable results. Wilson and Wales found that older participants, in general, were able to produce more complex compositions. Also, more girls than boys earned stage three ratings for complexity. These differences in complexity with respect to sex and age were more marked for rhythmic complexity than for melodic, a finding that is related to Hagen's (1996) finding that participants seemed to compose with higher composition validity for unpitched instruments. The ability of participants in both studies to develop compositions melodically seemed to progress more slowly than their ability to develop the timbral or rhythmic elements of their compositions. Wilson and Wales also reported that most students preferred to compose sequentially, from left to right, instead of a more random approach. Participants who used the playback function more frequently tended to score lower in melodic complexity, perhaps echoing the findings of Kratus (1994) and Hagen (1996) that higher aptitude (and likely compositional validity) corresponded with less exploration. It seems likely that participants who needed to hear their compositions played back would have had lower audiation ability, but Wilson and Wales did not provide music aptitude data. Han (2004) found that many professional composers cautioned against overuse of the playback functions of software because they believed it shut off the inner ear – or at least halted its development. Wilson and Wales also observed that less than one third of the participants were observed to use the eraser function. This could be related to Strand's (2006) claim that revision and reflection are often overlooked in music composition pedagogy.

Folkestad, Hargreaves, and Lindstrom (1998) explored the processes students use to compose by investigating the strategies used by young composers. In their study, 15- and 16-year-old participants (n=14) composed music using computer software over a three year period. The researchers systematically collected and recorded the compositions, made observations, and interviewed the participants.

Folkestad, Hargreaves, and Lindstrom (1998) identified two general strategies. First, horizontal composers tended to create the composition one horizontal line at a time. Then, during subsequent composing sessions, other lines would be added to fill out the texture. According to the researchers, the compositions of these composers were first “completed in form and context from beginning to end” (p. 88) and then arranged to their final form. This strategy of composing matches that observed by Wilson and Wales (1995), who found that students tended to prefer composing from left to right. For horizontal composers, composing and arranging were separate processes. Second, Folkestad, Hargreaves, and Lindstrom described vertical composers. In these pieces, “each section of the composition [was] completed for all instruments before moving on to the next” (p. 90). Vertical composers composed and arranged at the same time. Based on these qualitative findings, the researchers claimed that students use a variety of strategies when composing, and that there is no one ‘right’ way to compose.

Questions about the advantages, challenges, and values of using music technology in composition pedagogy were raised by Folkestad, Hargreaves, and Lindstrom (1998), Wilson and Wales (1995), Dammers (2007), and Prasso (1997). To provide the perspective of professional adult composers, a group who composes as a portion of their livelihood instead of as an activity in music class, Han (2004) interviewed established professional composers (n = 20). The purpose of the study was to ascertain whether there was a consensus about the advantages, disadvantages, and challenges associated with computers and composition. Four composers said they used computer music technology for all of their writing tasks, eleven said they used computer music technology sometimes, and five said they never used computer music technology. There were no commonly held beliefs among the composers; in fact their opinions differed drastically. Composers disagreed on the value of the immediate aural feedback that technology enables. First, some held the belief that the immediate aural feedback enabled composers to experience their music as a more objective listener, a finding similar to that of Prasso (1997). Second, it allowed composers to create music too complex or difficult for their own keyboard skills to realize. Third, it enabled composers to aurally identify performance problems and incorrect notes before rehearsals. Fourth, while no one interviewed claimed that computer music technology could substitute for an orchestra, several composers said that it allows young composers to experiment with the sounds of an orchestra in its absence. Other participants, however, viewed aural feedback as a disadvantage. One participant said that aural feedback was likely to prevent a composer from exercising his or her "inner hearing" (p. 138). This echoed concerns from other participants who warned against young composers using the software as a crutch to compensate for a lack of ear training. This opinion seems to verify Wilson's and Wales's (1995) finding that participants who create more complex compositions rely less on aural feedback and more on an internal model.

In conclusion, music educators have not reached consensus about the value or best pedagogy for teaching composition (Strand, 2006). Although many studies employed the use of computer technology in composition (Dammers, 2007; Hagen, 1996; Hickey, 1995; Kratus, 1989, 1994 and Prasso, 1997), there is substantial disagreement among established composers about the value of that technology (Han, 2004). Researchers have found that activities in music composition can increase students' musical achievement, specifically their instrumental music achievement (Gaynor, 2005) and sight-singing achievement (McPherson, 1994). At the very least, teaching composition in place of standard classroom activities does not negatively affect student performance, sight-reading, or music theory achievement (Guderian, 2008). It is also clear from the research (Folkestad, Hargreaves, and Lindstrom, 1998) that young composers employ a large variety of strategies when composing. This finding is perhaps most related to creativity studies (Hickey, 1995; Kratus, 1989, 1994; Webster, 1979) where substantial differences in the processes used and the characteristics of the musical products created have been found.

The division of this first portion of the literature review into three categories seems to mimic the division of researchers into one area or another, with a few exceptions (for instance Hickey, 1995 and Josuweit, 1991). Unfortunately, studies on creativity seem to only tangentially address composition issues like tonal pattern vocabulary and audiation ability. Likewise, researchers who have explored the learning of tonal patterns have seldom referenced what is known about creativity (Kratus, 1994; Stoltzfus, 2005).

In the present study, I addressed this problem in two ways. First, I employed aspects of all three categories. Participants were allowed to employ their creativity by creating patterns through improvisation. They were required to use their audiation skill when echoing patterns. They also completed an activity where they experienced the composition process. In this way, I was guided in the design of the present study by ideas and established theories from all three categories of research. Second, my study helped to fill a gap in the research: no study was found that identified how accurately students could audiate their compositions. Knowing whether fourth graders can actually audiate simple, self-composed melodies – their level of composition validity – can shed light on how they create music.

The Importance of Visual Stimuli

While music is primarily an aural art form, Barrett (1990) and Auh and Walker (1991) have claimed that visual stimuli play an important role in the comprehension, recall, and composition of music. Barrett found that a substantial proportion of participants engaged in a melody-reproduction problem relied on iconic representations of contour and pitch. Auh and Walker found that young composers who wrote their compositions in invented graphic notation instead of standard notation demonstrated a higher level of creativity and used a greater number of compositional strategies.

Barrett (1990) explored how fourth graders (n = 20) solved a melody-reproduction task. The participants were challenged to perform a familiar melody (for 16 participants, *Are You Sleeping?*; for four participants, *Row, Row, Row Your Boat*) on a xylophone in order to investigate their verbal and performed responses to the problem. Participants first completed a warm-up activity that reviewed how to perform skips and steps, and then attempted to reproduce the melody. After the reproduction task, participants were interviewed. While attempting to solve the problem, Barrett provided participants with various levels of assistance (scaffolding). Barrett (1990) found that participants solved the melodies by comparing their performance with stored constructs called schemata. Participants used *general melodic schemata* (representations of musical structure based on experience) to arrive at plausible answers. Participants also used *song-specific schemata* (the recall of specific interval or contour characteristics). Barrett found that participants used three basic strategies to solve the musical problem: *temporal order* (solving the melody one pitch or one phrase at a time, and then performing that part and solving the next), *comparison* (systematically trying one pitch at a time until the right pitch was found), and *inferring strategies* (making educated guesses based on experience). Barrett grouped students into three group types. First, *aural solvers* relied upon their inner hearing and used song-specific schemata and comparison strategies to reproduce the melody. Second, *visual solvers* relied upon iconic forms of representation (participants were instructed to represent the relative pitch and duration of phrases with specially-designed cards) and used comparison and inferring strategies. Third, one participant was unable to solve the given melody. A relationship between the group types and music aptitude might have provided a better understanding of why different participants performed as they did, but no such data were provided. Given the findings of Hagen (1996), who found a negative correlation between music aptitude and amount of exploration, it seems likely that high aptitude participants would have likely been aural solvers, medium aptitude visual solvers, and low aptitude non-solvers. The finding that several participants arrived at solutions that were correct in contour but not in interval (i.e. participants were able to perceive where melodies rose and fell, but the intervals used in their performances were not exactly correct) is further evidence that contour is a stronger factor in the recall of memories then the discrete pitches of the melody, a finding also reported by Dowling (1978). In summary, Barrett (1990) found that participants use a variety of melodic schemata and cognitive strategies to solve musical problems.

Auh and Walker (1991) analyzed the compositions of seventh graders in Korea (n = 38). Half of the participants were instructed to compose music using standard notation;

the remainder were allowed to create their compositions using invented graphic notation. The purpose of the study was to discover how the use of graphic notation affected creativity and the choice of compositional strategies. Auh and Walker found that the group that used graphic notation employed a significantly greater number ($p < .02$) of compositional strategies in the process of creating their compositions. The graphic notation group also showed significantly higher creativity ($p < .02$) on a researcher-designed creativity measure. These findings led Auh and Walker to claim that graphic notation leads to greater divergent thinking skills. Perhaps, in light of Webster's (1979) claim that fluency and flexibility lead to divergent thinking, the opportunity to create music without considering the bounds and limitations of standard notation allowed the participants greater flexibility. In conclusion, both Barrett's (1990) finding that some students solve musical problems by visualizing contours and pitches and Auh and Walker's (1991) finding that allowing students to compose in graphic notation support the claim that visual stimuli is important for comprehending, recalling, and composing music. Based on these findings, I chose to include a contour activity in the composition activity in the present study.

Contour and Memory for Melodies

Cooper (1994) investigated whether error-detection ability increases with age. Participants (N=169), ranging in age from grade one through grade six, were each given 30 melodic pairs to listen to. Each melody consisted of five pitches. Participants were instructed to determine sameness or difference between the two melodies, and determine which pitch was changed and in which direction if there was a difference. Each difference in pitch was smaller than a semitone in magnitude. Unlike Dowling's (1978) study, there were no differences in contour between each of the melodies in a pair.

Cooper (1994) found that participants of all ages were able to perceive differences between the melodies, a significant finding ($p < .05$). Further, participants were able to perceive differences even when they were unable to identify which pitch was different or in which direction it had been altered. Cooper found no significant difference in error-detection ability between consecutive grade levels, but there were several significant differences at the .05 level when the age difference exceeded two or three years (e.g. fifth graders were significantly more adept at the task [$p < .05$] than were first and second graders).

Saunders and Holahan (1993) further explored the abilities of participants to detect differences between paired melodies. Second graders (n = 46) completed the IMMA (Gordon, 1986) in both its standard format and a researcher-designed computer-based version. The researchers compared the participants' scores to determine the reliability and validity of the computerized measure and examined the response time for each item.

Saunders and Holahan (1993) reported several significant findings. First, high-scoring participants (i.e. participants with high aptitude) responded faster ($p < .05$) for correct judgments of sameness. Second, both high-scoring and low-scoring participants responded more slowly ($p < .05$) when making incorrect judgments of sameness. Third, high-scoring participants made correct different judgments faster ($p < .05$). Fourth, boys responded significantly ($p < .05$) faster than girls, although they did not respond more or less accurately.

Saunders and Holahan (1993) made several claims based on these findings. First, they claimed that because of the relative ease with which participants were able to determine sameness (and the smaller amount of data for incorrect same responses), less is known about same judgments than different judgments. To investigate this, they recommended developing tests with more difficult same items. Second, Saunders and Holahan claimed that their response time findings suggest that incorrect judgments of sameness are made not impulsively but after a substantial thought process. This confirms the finding of Cooper (1994) that difference can be determined without being able to identify how or where the difference occurred. Participants who were not able to quickly compare the two melodies may have had difficulty finding a difference. Third, Saunders and Holahan suggested that higher-scoring participants retain melodies more quickly, are better able to mentally represent those melodies, and are better able to compare and contrast those melodies. Finally, Saunders and Holahan called for more research about item difficulty on similar discrimination tests, including research on the effect of the intervallic content of test prompts. I applied such an analysis in my study by examining participant responses with respect to the intervallic content of the melodies compared by participants.

Holahan and Saunders (1997) continued their research by investigating the effect of contour on response time and item difficulty. Specifically, they sought to find out how similar and different contours affect the ability of participants to correctly determine sameness or difference. For the test, they used 45 three-pitch melodic pairs. Of these, 12 pair were identical, nine pair were different and had opposite contour, six pair were different but had the same contour, and 18 pair were different and had contours that were the same for one interval and different for the other. The computerized IMMA test used in previous research (Saunders and Holahan, 1993) was modified for this study. The response time was recorded for all participants.

Holahan and Saunders (1997) found that participant success was significantly related to contour type. Participants were divided into low, average, and high achievement groups based on their score. The low achievement group scored significantly less ($p < .001$) than the high achievement group for every contour type. The average achievement group also scored significantly lower ($p < .05$) than the high achievement group, but not on same-contour items. This finding corroborates the earlier (Saunders and Holahan, 1993) finding that sameness is easier to detect than difference. Holahan and Saunders (1997) reported that the test developed for the study was easier for participants than they had planned. Because Holahan and Saunders were able to generate several significant results by asking fourth grade students to complete discrimination tasks with three-pitch patterns, I decided to use three-pitch patterns in the composition activity in my study. Even though Holahan and Saunders reported that the patterns were too easy for their participants, I expected that the added cognitive burden of needing to hold a melody in memory would prevent the patterns in my study from being too easy, especially viewed in light of findings by Kratus (1989, 1994) that such memory function is quite challenging for students of this age.

In summary, there is substantial evidence behind the claim that contour is an important schema for remembering and comparing melodies (Dowling, 1978). Davies (1979) disputed the logic of Dowling's conclusions, suggesting that instead of contour melodies were perceived as holistic units. Cooper's (1994) finding that students are often unable to locate a difference even when they are able to detect one seems to agree with Davies's notion. Holahan and Saunders (1997) supported Dowling's idea of contour as a schema when they found that the sameness or difference of the contour of two patterns substantially affects the difficulty of making that determination. There will be a contour activity in the proposed study, if not for the reasons that Dowling, Saunders, and Holahan (1993) suggest than because of Barrett's (1990) classification of visual solvers. Thus, contour served a dual purpose in the present study: it acted as a visual cue for some participants, and it may have acted as a memory aid for others.

Conclusion

Based on the reviewed research, I found that participants' degree of composition validity is a product of many disparate factors. The age, sex, level of formal and informal music experience, music aptitude, creative potential, and intrinsic motivation of participants play important roles in determining how closely young composers' compositions resemble the composers' intentions. An assessment of whether audiation-based pedagogies for composition could be the solution to enabling composition validity is beyond the scope of my study; however, in my study I developed a foundation for addressing that possibility. I sought to quantify the composition validity of fourth graders. Once it is established that fourth graders are indeed able to audiate their compositions – a finding that before my study had not been reported in any available empirical studies – an examination of the promise of audiation-based composition can be conducted.

Chapter 3

Methodology

The purpose of this study was to explore the composition validity of fourth graders by evaluating whether they were able to audiate tonal patterns that they have composed. In addition, I investigated relationships between that ability and several pertinent variables, including participant music aptitude, sex, age, and school assignment.

Research Questions

I investigated four research questions. *(1) Are participants able to audiate tonal patterns that they have composed?* Gordon's Intermediate Measures of Music Audiation (IMMA) (1986) has been shown to reliably measure students' music aptitude by challenging participants to tell sameness or difference between two given tonal patterns (1984). In the present study, I challenged participants to perform the same task between one audiated tonal pattern and one given tonal pattern using the Discrimination of Sameness or Difference (DSD) measure.

(2) Is there a relationship between participants' ability to audiate given tonal patterns and their ability to audiate patterns that they have composed? McPherson (1994) found that audiation ability increases sight-singing achievement, and Stoltzfus (2005) reported that audiation-based composition increases performance achievement. Based on these findings, it is reasonable to expect that music aptitude is a reliable predictor of success on the DSD measure. To demonstrate this, scores on the DSD measure were correlated with scores on IMMA.

(3) What effect do age, sex, and school assignment have on the ability to audiate self-composed tonal patterns? Wilson and Wales (1995) found that second and fourth grade girls were able to compose more complex compositions than boys, and Saunders and Holahan (1993) found that second grade boys responded significantly faster than girls of the same age. I chose to control for sex in my study to determine if the ability of girls to compose more complex compositions or the impulsivity of boys was a significant factor. Kratus (1989) found that older students used less exploration, and Wilson and Wales found that the complexity of participants' compositions increased with age. I chose to control for participant age in my study to determine if older students were better able to audiate the simple compositions employed. Another reason for controlling for age was to prevent participant maturity from confounding my findings. Participants in my study were assigned to their elementary schools based on their home address. While both schools are in the same city and school district, the two schools where the study was conducted operate under different administrators with different priorities, scheduling practices, and philosophies. For these reasons, controlling for age, sex, and school assignment was necessary.

(4) What effect do intervallic content (such as the presence or absence of unisons, seconds, and thirds) and response type have on the ability to audiate self-composed tonal patterns? A sample of the patterns composed by participants was coded categorically according to their interval content. The number of correct responses for each category of interval content was used to determine whether particular attributes made patterns easier or more difficult for participants to audiate. Similarly, the number of correct responses

for each category of response type (whether the two patterns compared by participants were the same, different in the second pitch, or different in the third pitch) was used to determine whether response type affected difficulty level.

Research Design

The present study was a quantitative, causal-comparative study. The decision to model the proposed study after similar psychological studies on error-detection and judgment of sameness or difference is supported by Holahan and Saunders (1997), who wrote that "the mental processes underlying audiation are not fully understood, but they may be most effectively elucidated by systematic research in musical and cognitive psychology frameworks" (p. 87).

Participants

The present study was conducted among fourth graders (N = 152) in two elementary schools in an urban school district in Massachusetts. Some students in the fourth grade population did not participate in the study due to substantial difficulties with English proficiency or documented learning disabilities that would have prevented them from completing the tasks. The fourth grade was chosen because of Gordon's claim (1986) that music aptitude has stabilized and the precedent for composition studies at this age (Barrett, 1990; Josuweit, 1991; Wilson and Wales, 1995).

I chose to conduct research with fourth graders for several reasons. First, it has been my experience that students of this age are able to compose simple melodies with varying degrees of composition validity. Second, Gordon's (1986) claim that music aptitude has stabilized around this age reduces the validity threat of maturation –

participants may be more or less likely to succeed based on high or low music aptitude, but not based on age. Third, there is a precedent in the research for conducting similar studies with students of this age (see Barrett, 1990; Josuweit, 1991; Wilson and Wales, 1990). Fourth, the participants have been prepared for composition activities of this difficulty level through the music curriculum and lessons that I have developed.

The population served by the school district is generally more racially diverse and has a lower socio-economic status than the state average. During the 2009-2010 school year,[4] 45.1% of students were White, 40.0% were Hispanic, and 14.9% represented Asian, African American, Native American, or Multi-Race families. Nearly half (44.7%) of students spoke a first language other than English. As a comparison, the state average was 69.1% White, 14.8% Hispanic, and 16.1% other races during the same year. Statewide, only 15.6% of students spoke a first language other than English. Of the students enrolled in the 2009-2010 school year, 57.3% received free lunch, and an additional 13.6% of students purchased reduced-price lunch (the district had 71.2% low-income students; the state average was 32.9% low-income students). The school district had a special education population of 15.0%, slightly below the state average of 17.0% (Massachusetts Department of Elementary and Secondary Education, 2010). The participants in the present study reflect the demographics of the district; however, they were not randomly assigned to school groups and so should not be considered a representative sample.

The present study was completed in two different school buildings. Although the

[4] This is the most recent year for which demographic data is publicly available.

two schools are less than a mile apart, two very different environments exist. Walking into School A, an outsider would first notice that there is no surplus space. The school is filled past capacity, and students are seen working on tables set up in closets and hallways. The music room is on a small stage (a mere 10 feet deep!) in the auditorium/cafeteria. Students sit on crowded choral risers (there are no chairs available) in class, and the stage must be rearranged often for instrumental music and drama performances. School B, by comparison, is much larger. Although the class sizes are nearly identical, the hallways are wide and clear and the rooms spacious. There is a well-appointed music room with chairs for all the students. Music rooms in both schools are equipped with a piano, several electronic keyboards, and an assortment of classroom instruments. I chose to conduct the present study in these two schools because I have been the general music teacher in the schools for approximately five years.

Sampling procedures and sample size.

All of the 165 students in fourth grade were invited to participate in the present study. Parents of participants were notified by letter and offered an opportunity to have their children included in the study. First, all participants (N = 152) completed the IMMA (Gordon, 1986) in whole-class settings during their regular general music period. Students who had been excluded from the study still participated in this first whole-group session to avoid singling them out. Their scores were later discarded. Second, I used a random number table to choose participants from each school to complete the DSD in individual sessions with the researcher (n = 64). I chose 32 participants from each school for this sample to ensure that there would be enough participants in each subgroup for age, sex, and school assignment. From School A, 29 students completed the DSD, 17 boys and 12 girls. From School B, 31 students completed the DSD, 16 boys and 15 girls. Four participants were chosen to be in the DSD sample but did not participate because there was not enough time during the two days of sessions. Third, a stratified random sample was used to identify participants from the DSD sample whose composed melodies would be used for further analysis of intervallic content (IC). Three boys and three girls were chosen from each school (n = 12) for this analysis.

Ethical treatment of human subjects.

Steps were taken to ensure that participants were treated ethically during the study. Because the present study took place in a public school by me (a licensed music teacher in Massachusetts) and the participants' activities (e.g. composing, listening, responding, etc.) were not substantially different than activities normally conducted in general music class, I assumed that the study involved not greater than minimal risk for participants. Participants gave verbal assent, and their parents were informed of the study before it began and given an opportunity to opt out. In addition, I gained consent from building principals at both research sites and the district Fine Arts Director. All participants who completed the IMMA were given a gift of a new, colorful, music-themed pencil. All research was approved by the IRB before the study began. At all times during the present study, I conformed to the standards of Boston University's Institutional Review Board (IRB).

Data Collection

I modeled the data collection process of the present study after related studies.

The procedure closely resembles the IMMA (Gordon, 1986). The composition task, which participants completed by manipulating large cardboard notes with iconic representations of pitch, was modeled after the research of Barrett (1990).

Instruments.

The IMMA (Gordon, 1986) was administered in group settings to approximately 30 participants at a time. The directions in the test manual relating to giving instructions, conducting the test, and scoring the tests were followed explicitly. Only the tonal subtest was given. Gordon (1984) supported the reliability of the IMMA by performing a split-halves analysis (r = .72) among fourth grade students. In addition, Gordon performed a longitudinal study and determined that the measure was a valid predictor of children's later music achievement.

Participants completed the researcher-designed DSD measure that closely mimicked the IMMA (1986). The principal difference between the IMMA and the DSD is that participants compare two given patterns for sameness or difference on the IMMA; they compare one given pattern and one composed pattern on the DSD. To complete the DSD, the participants first used a series of cards to compose a melody (see Appendix E). Second, the researcher performed the composed melody. Third, the participants circled their choice for sameness or difference, comparing the melody they composed (but had not heard) with the melody the researcher performed.

Of the 10 items on the DSD, six melodies were played exactly as they were composed by participants. This response type is called *same*. Four were played with deliberate errors made by the researcher in performance. Two of these errors were made by changing the second pitch and two of these errors were made by changing the third pitch. These response types are called *different-2* and *different-3* respectively. The decision to play "same" response types for six of the 10 items was based on Saunders and Holahan's (1993) finding that sameness is easier to detect than difference. I chose to weight the measure in this way to prevent it from being too hard for participants, a possibility that seemed quite likely based on my own experience with students being unable to recognize their compositions.

Like the IMMA, the measure has four practice items, two of which have been completed for the participants. A heavy black line separates the practice items from the scored items. Also like the IMMA, test items are indicated by simple graphics (e.g. a book, a tree, etc.) instead of words. According to Gordon (1986), this representation is preferred to numbers or words for students with language or reading delays. The placement of the *same, different-2* and *different-3* items in the DSD were determined with a random number generator prior to the first session. The order was the same for all participants.

While this measure had not been tested for its reliability at the outset of the data collection, several steps were taken to ensure the data produced were as reliable as possible. First, I was in close proximity to participants while they were completing the measure, preventing them from circling an answer for the wrong item. Second, the same response items were played incorrectly for each participant. This lessened the possibility of researcher error during performance and made the use of a scoring mask possible. Third, I performed all prompts at a consistent tempo of 108 beats per minute, the tempo used in the IMMA test (1986). Fourth, I used an electronic tuner to check the three required pitches prior to and directly following each day of data collection. Intonation differences were less than 5 cents different from their 440 Hz. standard. Finally, the substantial sample size in this study helped to minimize the impact of any uncontrolled instrument reliability threats.

The measure had not been tested for validity at the outset of the data collection. There were several possible threats to the validity of this instrument. First, participants might have been able to see the researcher playing the test prompts and might have been able to use their knowledge of keyboard skills to determine errors. This threat was reduced by large cardboard barriers that were placed on the music rack of the keyboards. Second, validity was demonstrated through analyzing student responses. Overall, participants scored significantly better ($p < .0001$) than chance on their responses on the DSD, a finding that will be examined in greater detail in the next chapter. Further, a positive correlation ($r=.258$, $p < .05$) between scores on this measurement and IMMA (1986) scores provided additional evidence that the test measured the participants' ability to audiate their melodies. Because the DSD task required participants to audiate a melody, it follows logically that there is a positive relationship between the two scores.

Research procedures.

Participants began the study with a broad range of formal music experiences with tonal patterns and solfege. In first grade, this cohort of students learned to sing the major scale using solfege syllables and identify the major scale as it appeared in their classroom repertoire. In second grade, they learned to sing several songs through reading printed solfege syllables during an extensive unit on contour. For participants who arrived in the district as new students at the beginning of fourth grade, their understanding of solfege and tonal patterns was aided by weekly lessons echoing music with their soprano recorders (taught in a solfege-only approach). During the first week of the study, all fourth grade classes participated in a lesson (Appendix D) designed to review their previous learning of tonal patterns and solfege and to teach them how to complete the activities required in individual sessions.

During the second week of the study, participants completed the IMMA (Gordon, 1986) in whole-class sessions. The IMMA was administered during participants' regular music classes in the music room at each school. I administered the IMMA myself, explicitly following the instructions provided in the test manual. The audio CD of the test material was played through a Fender PD-250 sound system.

During the third and final week of the study, participants in the DSD sample completed the DSD measure in individual sessions with me. These sessions were scheduled throughout 2 days (one in each school). Each session lasted approximately 10 minutes. As each participant completed their session, they were asked to send the next participant to the room. For the DSD sessions, a small, quiet area was equipped with chairs for me and one participant at a time, a small table, a Yamaha YPT-220 keyboard, a Korg electronic tuner, a Logitech webcam, and 12-inch cardboard partitions to prevent the participant from seeing my performance of their melodies. A deck of specially-designed cards, 3 inches wide and 6 inches tall, was used to limit the pitch choice of participants to *do, re*, or *mi* (see Appendix E). Cards were printed so that *mi* appeared at the top of the card, *re* in the middle, or *do* at the bottom to aid students in visualizing contour. Fifteen cards of each pitch were provided, but not all of these were used in each session. At the beginning of each session, three of these cards were placed within the reach of the participant reading *do-re-mi* from left to right. All DSD sessions were recorded; the camera's perspective showed the participant workspace. To ensure consistency, the researcher conducted each session according to the script that follows.

> Today we're going to work together making up songs and playing a little game. First, look at this little song I wrote. It has three notes, and it goes like this. [*Researcher plays do-re-mi on the keyboard.*] While I play it again, follow the contour with your finger like this. [*Researcher demonstrates.*] Be sure to touch each note. Are you ready to follow the contour with your finger? [*Researcher waits for response, and then plays do-re-mi while student follows. Before continuing, researcher checks for understanding (And explains again and repeats the task if necessary).*] Did the notes in the song you heard go up? [*Student responds and then researcher gives feedback "that's right, the notes in the song went up," or "listen more closely, the notes in the song went up."*] Does the contour you see go up or down? [*Student responds and then researcher gives feedback as before.*] Look at the Cup Song on your paper. Because the notes you heard are the same as the contour you see, the two faces that are the same in the top box are circled.
>
> Let's do another example. This time I'll change our song a little bit. [*Researcher adds cards on top of those already on the table so that the notes read mi-re-do.*][5] While I play this song, follow the contour with your finger and touch each note. [*Researcher plays mi-re-<u>mi</u> while student follows.*] Did the notes in the song you heard go down? [*Student responds and then researcher gives feedback.*] Does the contour you see go down? [*Student responds and then researcher gives feedback.*] Look at the Tree Song on your paper. Because the notes you heard are not the same as the contour you see, the two faces that are different in the bottom box are circled.
>
> This time, follow the contour as you listen to the song I play. If I play a song that sounds the same as the contour you see, circle the two faces for

[5] A sufficient number of cards were provided so that each melodic change was achieved by adding a new card on top of one of those already on the table. Each melody was three pitches long.

> the Pencil Song that are the same in the top box. If I play a song that sounds different than the contour you see, circle the two faces that are different in the bottom box. [*Researcher adds a card so that the notes read mi-do-do.*] Are you ready to trace the contour with your finger, listen to my song, and circle the box? [*Researcher waits for response, and then plays mi-do-do while student follows. Student circles an answer, and then the researcher gives feedback.]*
>
> Let's do one more for practice. [*The cards remain the same.*] Are you ready to trace the contour with your finger, listen to my song, and circle the box for the Apple Song? [*Researcher waits for response, and then plays mi-do-re while student follows. Student circles an answer, and then the researcher gives feedback.]*
>
> Now we're ready to start our game. Each time, you will add one note to change your song, and then be ready to follow the contour, listen to the notes, and circle the box for that song. Go ahead and add one card to change your song.
>
> [*The following is read before each of the ten response items.]*
>
> Are you ready to hear the *x* Song? [*Student responds, researcher plays, student circles.*] Now choose a new note for the *y* song. [*Procedure repeats for all ten test items.*]

Following the tenth song, I thanked the participants and asked for their help returning the cards to the proper stack for the next participant. During the session, I helped participants to understand the instructions or directed participants to the correct test item as needed, but did not offer any other assistance, give any other feedback, or repeat any response items.

Data Analysis

Procedures for scoring tests.

IMMA tests were scored according to the procedures provided in the manual (Gordon, 1986). Scoring masks were used, and I rechecked 10% of the tests to ensure reliable scoring. The DSD tests were scored in a similar fashion. Once the incorrect items had been determined, a cardboard scoring mask similar to the one provided with the IMMA test was prepared. All tests were scored using this mask. I again rechecked 10% of the completed DSD tests for accuracy in scoring.

Procedures for collecting other data.

The interval content (IC) of a sample of the patterns composed by the participants (n = 120 patterns from 12 students) was determined with the use of a researcher-designed Excel spreadsheet. Following the DSD sessions, I viewed the videos of 12 participants, entering each pattern into a cell using this syntax: *mrd*. A series of formulas was used to code the intervallic content of each pattern. The patterns were divided into six mutually exclusive categories: patterns containing *(a)* thirds only (like *dmd*), *(b)* seconds only (like *drd*), *(c)* unisons only (like *rrr*), *(d)* one second and one third (like *dmr*), *(e)* one second and one unison (like *ddr*), *(f)* one third and one unison (like *dmm*). I used these data, combined with the participant responses, to determine whether certain intervallic combinations were more difficult than others.

To protect the identity of participants, I coded their birthdates (obtained from school records) into one of three mutually-exclusive categories: *(a)* all participants younger than nine years, six months, *(b)* participants older than 9 years 6 months but younger than 10 years, and *(c)* participants ten years and older. Participants were also grouped by school assignment (School A or School B) and sex.

Software used in analysis.

For preliminary data coding (such as coding participant birthdates into age groups and coding intervallic content), Microsoft Excel spreadsheets were used. All other statistical analysis was completed using JMP software, version 8.0.2.

Research Validity

Several threats to internal validity were possible in the present study. First, participants come to school each day with a wide range of musical experiences and abilities. The research was undertaken in the beginning half of the year, before the fourth grade band program is offered (this begins in January each year). Second, some students likely performed differently on their individual session than they did in the whole-group session (when the IMMA was administered) due to anxiety. While this in itself is unavoidable, the individual sessions and assent language were called a “game” to help reduce participant stress or nervousness as much as possible. While some participants did appear nervous upon arriving at their session, they seemed to me to be very comfortable performing the tasks once they had had an opportunity to practice. Also, students were reminded that their names were kept confidential and that the study was “not a test.” Third, the elapsed timeframe of the study was approximately three weeks. This eliminated the threat of mortality; no participant left the study. The elapsed time between IMMA test and DSD session for all participants was less than six days.

There are substantial limits to the external validity of the proposed study. The primary objective in conducting this research was to understand how the students in the district studied compose and audiate. While the findings may be of interest to other educators in other districts, the results are only generalizable to populations of a similar demographic and socioeconomic makeup.

Chapter 4

Results

In this chapter I report findings regarding to the reliability of the DSD instrument, as well as the findings regarding the data collected through IMMA testing and DSD sessions. These results address the research questions already discussed. *(1) Are participants able to audiate tonal patterns that they have composed? (2) Is there a relationship between participants' abilities to audiate given tonal patterns and their abilities to audiate patterns that they have composed? (3) What effect do age, sex, and school assignment have on the ability to audiate self-composed tonal patterns?(4) What effect do intervallic content and response type have on the ability to audiate self-composed tonal patterns?*

DSD Instrument Reliability

Before DSD results were interpreted, I confirmed that the instrument performed reliably in this study using the split-halves method. First, I compared the number of correct responses on the first 5 items to those for the second 5 items using a two-tailed t-test. Second, I compared the number of correct responses on odd-numbered and even-numbered items using a two-tailed t-test. There were no significant differences among the halves of the instrument at the .05 level (Table 1[6]). Evidence of instrument validity will be demonstrated in the analysis of Research Question 1.

---

[6] The same number of "same," "different-2," and "different-3" response types appeared in each of the halves reported in Table 1.

Table 1 - DSD Split-Halves Check

Response Items	n	mean	sd	*p*
First 5	300	4.03	1.02	.58
Last 5	300	3.97	1.09	
Odd	300	4.08	1.10	.17
Even	300	3.92	0.92	

Note. Mean denotes mean number of correct responses out of five total responses, and *p* value is the result of comparing each pair of halves using a two-tailed t-test.

Research Question 1

I answered the first research question *(are participants able to audiate tonal patterns that they have composed?)* through analysis of responses on the DSD measure (Table 2).

Table 2 - DSD Score

n	mean	SD
60	8.10	1.65

Note. Mean denotes mean score out of 10 total responses.

Because I found the mean DSD score for all participants to be significantly different ($p < .001$) than a chance score of 5 using a two-tailed t-test, the alternative hypothesis was accepted: participants' responses on the DSD measure were significantly different than if they had been chance responses. The answer to the first research question, then, is: Yes, participants were able to audiate tonal patterns that they have composed. This result also helps to establish the validity of the DSD instrument: if participants were able to answer correctly more than 80% of the time, then they were likely to have been audiating the patterns they composed.

Research Question 2

Before answering the second research question *(is there a relationship between participants' abilities to audiate given tonal patterns and their abilities to audiate tonal patterns that they have composed?)*, I examined to what degree the independent variables age, sex, and school assignment affected participant IMMA (Table 3) and DSD (Table 2) scores. Using MANOVA (Table 4), I found that school assignment significantly affected participant success (p = .047). This finding will be discussed with my analysis of the third research question.

Table 3 - IMMA Scores

Group	n	mean	sd
DSD sample	60	34.92	2.92
All participants	152	34.56	2.39

Note. Mean denotes mean score out of 40 possible correct responses.

Table 4 - MANOVA of IMMA, DSD Scores by Age, Sex, and School

Source	df	F	*p*
Age	2,48	0.47	.95
Sex	1,48	1.30	.26
School	1,48	4.17	.047*
Age x Sex	2,48	0.62	.54
Age x School	2,48	0.44	.65
Sex x School	1,48	0.20	.66
Age x Sex x School	2,48	1.28	.29

* $p < .05$

It should be noted that these data only partially met the assumptions necessary for MANOVA. First, I met the assumption of cell size due to the substantial size of the DSD sample. Second, the assumption of dependence was met by explicitly following the provided IMMA (Gordon, 1986) instructions, and by modeling the DSD instrument after the IMMA. Third, I verified the assumption of linearity between the dependent variables through analysis of a pairwise correlation between IMMA score and DSD score. Fourth, I analyzed the homogeneity of variance for all variables using Levene's test. Among all independent variables, only the difference in variance in IMMA score by sex was significant ($p = .037$). Fifth, I examined skewness (presented in chapter 5). While there was some deviation from the normal curve, I decided that these data were adequately normal given the substantial sample size. In summary, the assumptions of cell size, dependence, and linearity were met for these data. The assumption of normality was somewhat met, and the assumption of homogeneity of variance was met in all but one instance. Therefore, I conclude that while the use of MANOVA was appropriate, the results of the analysis should be interpreted cautiously.

Once I had assessed the effect of age, sex, and school assignment, I used a pairwise correlation to determine the strength of the relationship between participants' music aptitude and DSD scores. There was a correlation of .258 ($p < .05$), demonstrating a moderately weak, significant positive relationship. The alternative hypothesis for this question was accepted: a significant correlation existed between participants' DSD scores and their music aptitude. The answer to the second research question, then, is: Yes, there is a relationship between participants' abilities to audiate given tonal patterns and their abilities to audiate composed tonal patterns, but the relationship is relatively weak and may be confounded by school assignment.

Research Question 3

To answer the third research question *(what effect do age, sex, and school assignment have on the ability to audiate self-composed tonal patterns?),* I reviewed the MANOVA findings from the second research question (Table 4). Age and sex were not significant effects on participant success. School assignment appeared significant ($p <$ .05), but due to the data only marginally meeting the underlying assumptions for MANOVA such findings must be treated cautiously. To determine whether school assignment actually played a role in DSD score, I used a two-tailed t-test to compare the scores of participants from each school. Students from School A scored higher than those from School B, but the difference only approached significance ($p = .07$). Despite previous findings of the MANOVA that appeared significant, I chose to confirm the null hypothesis: age, sex, and school assignment did not significantly affect DSD scores. In summary, I made the decision to suppress my MANOVA findings and confirm the null hypothesis for two reasons: first, the assumptions for MANOVA were not soundly met, increasing the likelihood of false positive findings; and second, a two-tailed t-test revealed a non-significant difference.

Research Question 4

The fourth research question is *What effect do intervallic content and response type have on the ability to audiate self-composed tonal patterns?* I began by examining intervallic content (IC) type. I used a random number table to select 12 participants in the DSD sample. To ensure that the 12 participants chosen for this analysis were representative of the DSD sample, I compared their scores with the scores of the DSD sample using a two-tailed t-test. The scores were not significantly different at the .05 level.

First, I investigated the effect of the independent variable IC type on the dependent variable correct responses. I coded the melodies composed by these participants (n = 120 melodies, 10 per participant) based on their IC type: patterns containing (a) thirds only (like dmd), (b) seconds only (like drd), (c) unisons only (like rrr), (d) one second and one third (like dmr), (e) one second and one unison (like ddr), (f) one third and one unison (like dmm). The IC types (Table 5) are listed in order of the highest number of correct responses.

Table 5 - Correct Responses by IC Type

IC type	n	correct responses	percentage	SD
third, unison	16	15	94%	0.25
thirds only	10	9	90%	0.32
unisons only	9	8	89%	0.33
second, third	24	20	83%	0.38
seconds only	32	25	78%	0.42
second, unison	29	22	76%	0.44
Total	120	99	83%	0.38

Then, I used a one-way analysis of variance to determine the effect of intervallic content on the number of correct responses (Table 6).

Table 6 – One-way ANOVA of Correct Responses by IC Type

Source	df	SS	MS	F	*p*
Between groups	5	0.49	0.10	0.66	.66
Within groups	114	16.84	0.15		
Total	119	17.33			

Although the ANOVA did not reveal significant differences, I did notice a pattern among the category types (see Table 6): melodies containing thirds seem to be easier for participants to audiate successfully. It should be noted that the data did not meet the requirement for homogeneity of variance.[7] To further investigate this pattern, I re-coded the melodies into two categories: those containing at least one third and those not containing thirds (Table 7).

Table 7 - Correct Responses by Presence of Thirds

Contains third?	n	correct responses	percentage
yes	50	44	88%
no	70	55	79%
Total	120	99	83%

Despite the higher percentage of correct responses for patterns containing at least one third, a second ANOVA determined that melodies containing thirds were not significantly easier than those without thirds (Table 8).

Table 8 - One-way ANOVA of Correct Responses by Presence of Thirds

Source	df	SS	MS	F	*p*
Between groups	1	0.10	0.10	0.71	.40
Within groups	118	17.22	0.15		
Total	119	17.33			

The data for this analysis did meet the assumption for homogeneity of variance.[8] Despite melodies containing thirds being answered correctly more frequently than melodies containing smaller intervals, the IC type of a melody did not significantly affect the number of correct responses.

[7] For Levene’s test of homogeneity of variance, the variance for IC type was significant ($p < .01$).
[8] For Levene’s test, the variance for presence of thirds was not significant ($p < .01$).

Second, I examined whether the independent variable response type (whether the melodies were performed correctly or not) affected the dependent variable correct responses. For each participant, 6 patterns were performed correctly (response type of *same*), and 4 patterns were performed with at least one pitch altered (response type of *different*) (Table 9).

Table 9 - Correct Responses by Response Type

Response type	n	correct responses	percentage
same	72	67	93%
different	48	32	67%
Total	120	99	83%

I used a one-way ANOVA to determine that participants were very significantly more likely to answer *same* type prompts correctly than *different* type prompts (Table 10).

Table 10 - One-way ANOVA of Correct Responses by Response Type

Source	df	SS	MS	F	*p*
Between groups	1	2.00	2.01	15.45	.0001*****
Within groups	118	15.32	0.15		
Total	119	17.33			

****p<.0005

The data for this analysis did not meet the assumption of homogeneity of variance. In fact, the variances were very significantly unequal ($p < .0001$). As a result, I chose not to consider this finding significant due to the increased likelihood of committing a Type I error. I re-coded the response types to discriminate between whether the alteration of the participants' melodies occurred in the second (response type of *different-2*) or third (response type of *different-3*) pitch (Table 11).

Table 11 - Correct Responses by Different Response Types

Response type	n	correct responses	percentage
different-3	24	19	79%
different-2	24	13	54%
Total	48	32	67%

Then, I performed an ANOVA (Table 12). I determined that, while melodies performed with alterations in the second pitch were answered correctly less frequently than when the alteration occurred in the third pitch, this difference only approached significance.

Table 12 - One-way ANOVA of Correct Responses by Different Response Types

Source	df	SS	MS	F	*p*
Between groups	1	0.75	0.75	3.48	.069
Within groups	46	9.92	0.22		
Total	47	10.67			

The data for this analysis also did not meet the assumption for homogeneity of variance.[9] Again, I chose to consider the finding as non-significant to prevent a Type I error. It can only be said that items on the DSD of response type *same* appeared to be more likely to be answered correctly. Participants seemed more likely to successfully notice a difference if an alteration was made in the third pitch of their melody than in the second

[9] For Levene's test of homogeneity of variance, the variance for IC type was significant ($p = .0018$).

pitch, but this difference was not statistically significant.

Summary

I began my analysis by using two split-halves checks to determine that the DSD instrument produced reliable results. Neither responses from odd/even halves nor first/second halves of the DSD instrument were significantly different. Also, I found that the participants randomly assigned to the DSD sample were a representative sample of the IMMA group, and that differences in IMMA score were not attributed to participant age, sex, or school assignment.

For research question 1, I found that participants were able to audiate their composed melodies. Participants were able to score significantly better than chance ($p < .001$). Because they were able to identify their melodies when performed and indicate when their melodies were performed incorrectly, it follows that they were likely to have been comparing the performed melody to an audiated version.

For research question 2, I found that there was a moderately weak, positive correlation of r = .258 between music aptitude and IMMA score among participants in the DSD sample ($p < .05$). I also submit this finding as evidence of the validity of the DSD instrument: it follows logically that participants with a high music aptitude should be better equipped to audiate a composed melody.

For research question 3, I found that neither age nor sex nor school assignment significantly affected the success of participants on the DSD instrument. It should be noted that participants from School A scored substantially higher than participants from School B, but this finding only approached significance.

For research question 4, despite the result that participants scored better on melodies that included the interval of a third in their intervallic content (IC), this difference was not significant. Due to unequal variances, I was unable to claim that participants were significantly more likely to correctly answer items of the response type *same* than *different*. I was also unable to claim that they were significantly more likely to identify alterations made in the third pitch of their melodies than the second pitch; however, there was certainly a trend towards these findings. I can only claim that participants seemed to score better for *same* response items, and that they seemed more adept at identifying differences when they occurred at the end of a melody, not the middle.

Chapter 5

Discussion of Results

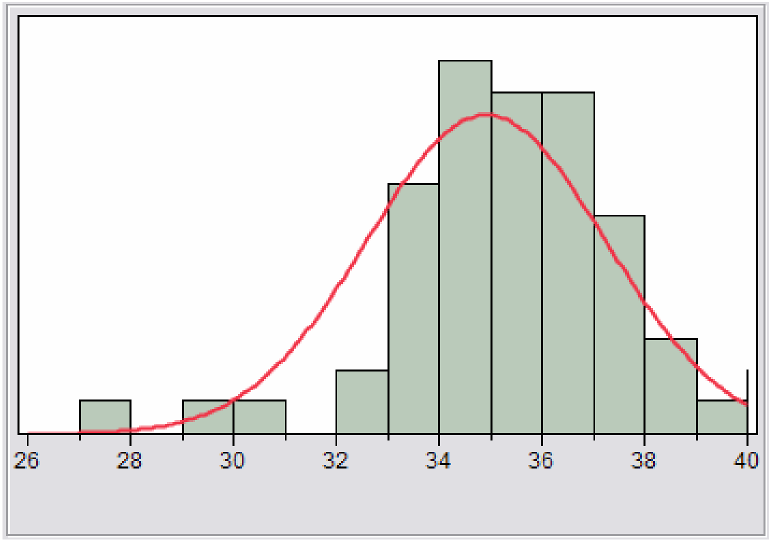
DSD Instrument Reliability

I confirmed the reliability of the DSD instrument using the split-halves method. Based on the lack of a significant difference in participant accuracy between either first half and second half or even- and odd-numbered prompts, participants were able to follow the instructions and complete the test without confusion. Based on the finding that participants earned a mean score of 81% on the DSD measure (this finding was significantly different than chance responses at the .001 level), it can be said with certainty that the participants were able to understand and apply the necessary skills to successfully complete the measure. Based on the participants' scores and the level of significance I conclude that the DSD instrument was an appropriate tool for assessing the ability of participants to audiate patterns they had composed.

Recommendations for future use of the DSD.

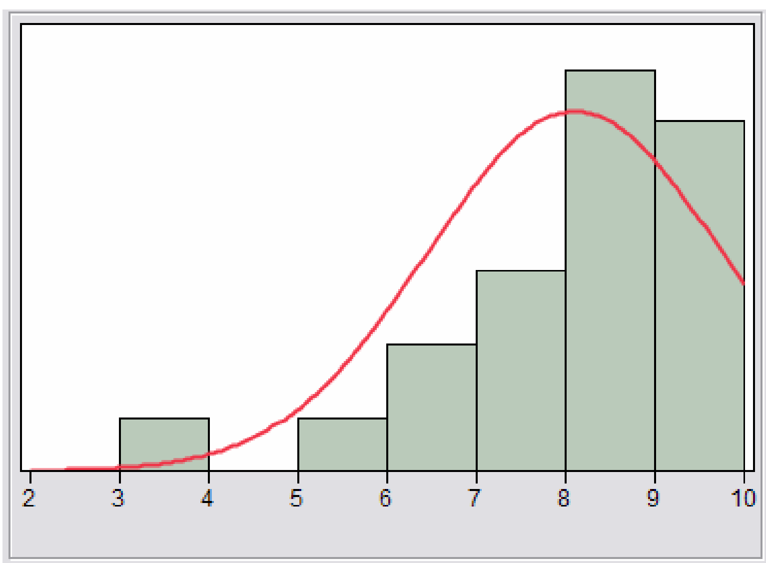
Both DSD scores and IMMA scores among participants in the DSD sample were negatively skewed, meaning that more participants scored above the mean than below the mean. It is clear through comparison of these DSD scores and IMMA scores that the DSD scores collected were more skewed than IMMA scores (see Figure 1 and Figure 2).

Figure 1 – Histogram of IMMA Scores



First, the skewness value of the IMMA scores of participants in the DSD sample (n=60) was -0.617. This distortion is relatively minor (the skewness value is approximately one-fourth of the standard deviation[10]).

Figure 2 – Histogram of DSD Scores



[10] Standard deviation was 2.324, and a skewness value of -0.617 represents .265 standard deviations.

Second, the skewness value of the DSD scores was -1.095. This distortion is also relatively minor (the skewness value is approximately two-thirds of the standard deviation[11]), but it is more skewed than IMMA scores. Had the DSD test been more difficult, it is likely that the skewness value of the DSD curve would have more closely resembled that of the IMMA curve; therefore, I suggest that future studies which rely on the DSD instrument should be weighted by increasing the number of the most difficult (*different-2*) response types and decreasing the number of the least difficult (*same*) response types (Table 13). The lower expected mean score would likely correct the histogram.

[11] Standard deviation was 1.654, and a skewness value of -1.095 represents .662 standard deviations.

Table 13 - Distribution of Response Types

Present Distribution

response type	items	mean score
same	6	5.59
different-3	2	1.58
different-2	2	1.08
total	10	8.25

Proposed Distribution

response type	items	mean score
same	2	1.86
different-3	2	1.58
different-2	6	3.25
total	10	6.70

Note. I computed the proposed distribution by applying the percentage of correct responses for each item type to the number of proposed items.[12]

Another possibility for correcting the histogram would be the inclusion of a greater number of pitches (i.e. do, re, mi, sol, and la instead of do, re, and mi) or the audiation of a longer melody (four or five pitches instead of three). In conclusion, based on the ability of participants in the present study to successfully complete DSD and observations about the distribution of responses, I made two implications about future use of the DSD

[12] For example, the percentage of correct responses for response type *same* was 93.1%. Therefore, the expected mean score for that response type for the 2 *same* items in the proposed distribution is 2 x .93, or 1.86.

instrument. First, the DSD instrument is a valid and reliable method for assessing the ability of participants to audiate melodies they have composed. Second, further research should be conducted to determine if a more difficult test would provide even more reliable results that more closely mimic the expected normal curve. This greater level of difficulty could be achieved with an adjustment in the distribution of item response type, a greater palette of available pitch choices, or a longer melody for participants to audiate.

The Ability to Audiate Composed Melodies (Research Question 1)

Participant scores were significantly higher than chance responses ($p < .001$), I conclude that fourth grade students are quite able to successfully audiate three-pitch patterns of their own compositions. Further evidence that the participants were successfully audiating is anecdotal. The following situation happened at least several times: A *different* response item was performed, the participants traced (correctly) the precise contour that was performed, and then marked *different* on their papers after seeing that the contour did not match the melody they had heard and were audiating. While no data were collected to confirm this occurrence, it remains that several students were not only able to identify that there was an error, but they were also able to tell exactly where and in which direction the error was made. It should be noted that being able to identify the direction and/or location of a melodic difference is not a prerequisite for determining that two melodies are different – indeed Holahan and Saunders (1997) found this to be the case. Anecdotal evidence such as this mirrors the findings of Stoltzfus (2005), who found that participants described audiation (albeit in their own words) when interviewed about how they compose. The finding that participants were able to complete the DSD is related to that of Saunders and Holahan (1993), who designed a similar test where participants were asked to detect sameness and difference between paired, three-pitch melodies. Participants completed the task so successfully that, like the DSD instrument, Saunders and Holahan called their test too easy.[13] In light of the findings of the present study and its similarity to related studies, I conclude that fourth grade students are able to audiate simple melodies that they have composed.

Based on that conclusion, I offer the following implications about the pedagogy of music composition. It should be noted that these implications are only valid for populations and samples of similar size and demographic characteristics as the participants in the present study. First, composition instruction that begins at a young age with the exploration of tonal and rhythmic patterns (see Dalby and Gordon, 2006), an approach often termed audiation-based composition, is the best approach to teaching students to compose. While the use of other approaches such as notation-based composition were not evaluated in the present study, these seem less likely to lead students to be able to create music with a high degree of composition validity.

Second, steps should be taken to ensure that students spend at least a portion of their instructional time creating compositions that employ a limited length and palette of pitches. This is not to suggest that students be restricted from creating freer compositions with an unlimited pitch palette, nor should it suggest that free compositions composed using student-invented graphical notation like that described by Auh and Walker (1991).

---

[13] In the present study, keeping the DSD instrument easy was a deliberate choice. Because no research studies were found that measured composition validity, I erred towards too easy instead of too difficult in order to increase the likelihood that the instrument would produce significant results.

Indeed, such a restriction would be likely to remove any creativity from the process. I only suggest that some of students' composition activities be structured in this manner to help them develop their composition validity.

Third, the assessment of student compositions should be connected to composition validity instead of craftsmanship or creativity. As Han (2004) found, many successful adult composers agreed with the suggestion that the development of a composers' inner ear should be their most important pedagogical objective.

The Relationship between Music Aptitude and DSD Score (Research Question 2)

There was a moderately weak, positive correlation of r = .258 ($p < .05$) between IMMA score and DSD score for students in the DSD sample. Several researchers whose studies I reviewed reported similar results. First, Auh (1995) found that music aptitude was not a significant factor in the variance of composition achievement, underscoring the weakness of the correlation found in the present study. I interpret Auh's finding as a caution to not overstate the presence of a positive correlation. Second, Kratus (1994) found that low music aptitude scores correlated with less time spent exploring, while high music aptitude scores correlated with more time spent developing. Kratus claimed that this difference was largely due to participants with a higher music aptitude being better able to retain musical patterns during a task. This finding could help explain some of the correlation found in the present study: participants who had a high music aptitude were better able to retain one pattern while comparing it to another. Third, Hagen (1996) claimed that audiation plays a role in the creation of new musical products. Based on the findings of my study and related findings by other researchers, I conclude that the ability to audiate is an important skill for the development of composition validity in young composers. Based on the weakness of the correlation, music aptitude is not a reliable predictor of composition validity.

Based on these conclusions, I made an implication about the ethics of teaching composition. The finding of a weak correlation supports Gordon's (1986) claim that the IMMA should never be used to exclude students from music activities. Indeed, among the participants who scored a "perfect" 10 on their DSD measure, approximately 30% scored well below the 35th percentile on their IMMA test. Had those students been excluded from a composition activity in the general music classroom based on their IMMA score, they might well have gotten the impression that they weren't good at composing – which certainly wasn't the case! Hickey's (1995) claim that music teachers were unlikely to be able to recognize their students' creative potential is also in line with this idea.

This finding further supports my preference for audiation-based composition. Students who are able to audiate well are somewhat more likely to be able to create compositions with a high degree of composition validity.

The Relationships among Age, Sex, School Assignment, and DSD Score (Research Question 3)

I found that none of these independent variables significantly affected DSD score at the .05 level. I had expected several of these variables to affect DSD performance based on similar findings in the reviewed research. As relating to age, for example, Cooper (1994) found that the ability to perceive pitch changes in melodies increases with age. Similarly, Kratus (1989) claimed that older participants were more likely to be able to retain patterns in their memories. My finding of non-significance makes sense when it is considered that both Cooper and Kratus used a wider range of participant ages in their studies.

I also expected to find a significant effect in DSD scores due to participant sex. Specifically, I expected girls to outperform boys. Saunders and Holahan (1993) found that males exhibited a faster response time. I had expected the boys' comparative impulsivity would lead them to complete each prompt before they had fully contemplated the two melodies. Further, Wilson and Wales (1995) found that females were capable of composing and performing more complicated melodies than males. This was not the case. Perhaps the boys' possible impulsivity enabled them to compare the melodies before forgetting them, or perhaps the substantial difference in the method of Wilson and Wales' study means that their findings should not be generalized to mine. Another possibility is that of a teacher effect: teachers in both School A and School B are predominantly female, and the teacher effect of me being a male role model was not controlled for in the present study.

I had expected based on the similar student populations, identical curriculum, and same music teacher that there would be no effect on DSD performance based on school assignment. Of the three independent variables addressed in this research question, this one most closely approached significance (p = .07). Further perplexing to me was the finding that students in School A – the one with the tiny music room and crowded choral risers – outperformed students in School B – the one with the well-appointed, spacious music room. Because none of the reviewed studies addressed school assignment, the following possibilities for this difference are largely conjecture. First, the cohort of fourth graders who participated in this study from School A received general music instruction for 30 minutes per week during their kindergarten year, where students from School B did not receive general music instruction until the beginning of their first grade year. Second, each of the two schools operates under a different principal with different philosophies, policies, and environmental challenges. This could introduce many uncontrollable variables (e.g. the level of ambient noise, participants' attitudes towards music, participants' anxiety level in test-like situations). Third, and most disheartening, though, is a personal observation that the so-called "specialists" (music, library, instructional technology, and physical education teachers) in School A are treated differently than at School B. Specialists are, in my opinion (which is likely biased), treated as a lesser status teacher at School B by administrators. At School B, they are assigned a greater amount of non-teaching supervision duties, included in fewer extra-curricular programs and professional development opportunities, and given a lesser amount of common planning time. In my opinion, colleagues at School A treat specialists as equal stakeholders who enrich the lives of students by broadening their experience. My perception is that at School B, colleagues view specialists like lesser teachers who serve them by giving them 30 minutes of preparatory time once each week. Related to the findings of Strand (2005), who found that a positive learning environment is crucial to developing students' composition abilities, it could be said that the more negative attitude towards specialists contributes to this difference. Based on these findings, and those of the reviewed literature, I conclude that the age and sex of students does not affect their ability to audiate their composed melodies; however, it has been my experience that school environment (specifically teacher morale) may impact student achievement.

The Effect of Response Type and Intervallic Content on DSD Score (Research Question 4)

The finding that participants seemed more likely to answer *same* items correctly than *different-2* or *different-3* is in line with the findings of Saunders and Holahan (1993). This makes the most sense when viewed from the perspective of the young composer. Because the composers who participated in the present study were able to identify differences in their compositions only about two-thirds of the time, I suggest that in general, young composers are likely to overlook mistakes in performances of their work about a third of the time. When these mistakes were in the middle of a tonal pattern, they seemed more even likely to overlook those mistakes. Based on the finding that tonal patterns containing thirds received some of the high percentages of correct responses out of all patterns, I suggest that these patterns contribute to a more extreme overall contour that may have made melodic differences easier to identify. Kratus (1994) found that participants with a low music aptitude were more likely to compose music that incorporated a broad pitch range. Perhaps this is another indication that broader-ranging contours are easier for participants (of all aptitudes, but specifically low aptitudes) to comprehend. This suggestion is supported by the complementary finding that patterns using seconds only or seconds and unisons in my study received a substantially lower[14] number of correct responses. Where the patterns including the skip to a third are easier for error detection, patterns built only with seconds require a more precise hearing to identify the same errors. It is also relevant that the standard deviations tend to increase with the difficulty level of IC types (Table 5). Not only are the patterns more difficult to comprehend, but there is also a wider dispersion of answers on these items.

Based on these conclusions and complementary findings among the reviewed research, I made two implications for the pedagogy if music education. First, because melodies that contained skips were easier for participants to audiate than those containing steps, I suggest that patterns with skips would serve better as introductory material. Patterns containing steps should then be introduced later, as students have developed and honed their abilities. Second, to aid students in increasing their ability to discern sameness or difference between their compositions and performances of their compositions, opportunities to perform their compositions or hear them performed by peers or teachers should be considered as an essential component of composition pedagogy.

Recommendations for Future Research

Several areas of related research could build upon the foundation of the present study. Experimental studies modeled after the present study would help to explain the importance and role of contour, music notation, and age in the audiation of young composers' creations.

[14] (but not significantly lower)

First, researchers could develop an experimental study to investigate the importance of contour. An experimental group would complete the study as described in this thesis, while a control group would complete a slightly different procedure. The control group's cards would have the dot for each note centered vertically on the card instead of raised or lowered relative to its pitch (see Appendix E). Also, the control group would not be asked to trace the contour with their fingers. Findings from this study could help to confirm the findings of several reviewed researchers. For example, Barrett (1990) found that participants in a melody-reproduction task relied on visual stimuli, while Auh and Walker (1991) found that participants who used graphic notation instead of standard notation composed more creatively.

Second, the procedures of this study could be completed by participants with the aid of a computer program similar to Saunders's and Holahan's (1993) adaptation of the IMMA test. In this study, a control group would be allowed a period of practice with auditory feedback (i.e. they would be able to hear their patterns through headphones as they compose) while the experimental group has no feedback. This study would help educators to understand the effect of "playback upon entry" feature in many popular music notation programs. This is related to the work of Prasso (1997), who found that sight-singing achievement was significantly better among students who composed with such auditory feedback than those who composed with the traditional pencil-and-paper method. Perhaps auditory feedback contributes to the abilities of students to comprehend their compositions.

Third, the study in its present form could be extended to provide a wider perspective on how children compose. For example, participants could complete response items for longer melodies or melodies with a larger pitch palette. Further, a larger number of participants representing a variety of age groups would help educators to determine an appropriate length for composition projects by their students.

Fourth, the remaining questions about the effect of age on DSD score – and especially why the middle age group outscored both older and younger peers – could be investigated with a cross-sectional or longitudinal study. A wider rage of age groups could be researched, possibly following the sampling procedure used by Cooper (1994) or Kratus (1994).

In conclusion, I hope that the results discussed in this study will be of value to other educators and their students. Perhaps more importantly, I hope that the present this study will be able to serve as a foundation for future research on this topic, enabling future generations of students to compose compositions that are as authentic to the artistic visions of their composers as possible.

Appendices

Appendix A – Verbal Assent Script

I want to tell you about something we are doing called a research study. A research study is when some people called researchers collect a lot of information to learn more about something. A research study may be like a science experiment or collecting information to solve a mystery. We are doing this study to learn more about how fourth graders write songs.

I am going to tell you about a study we are doing and then you can tell me if you want to do it or not. In this study, you and I will play a short song-writing game. You will write songs, I will play them for you on a keyboard, and then you will tell me if I played them right or wrong.

There may be some parts of this study that you may not like. You will have to miss about ten minutes of class. There is a little chance that there might be some other things that I don't know about right now.

You might find out that you're really good at writing songs if you do this research, but you also might not like writing songs with me. Even if you don't like writing songs, if you do this it could help other children by helping their teachers understand how they write songs.

If you say yes and do the research we will give you a special music prize.

There is a small chance that people, including your parents, may find out some private information about you that you tell us.

Are you wondering if you have to be in this study? No, you don't. No one will make you if you don’t want to do this. All you have to do is tell me if you don’t want to do it. No one will be mad at you or change how they take care of you because you don’t want to do it.

If you say yes and then later you change your mind it is ok, you can stop at any time. If you say yes and then someone asks you some questions you don’t want to answer it is ok for you to say no.

If you have any questions you can ask me right now. If you think of any questions later you can ask your teacher and they will ask me for you.

I will give you a copy of this paper if you want.

Appendix B – Parent Letter

Dear Parents/Guardians

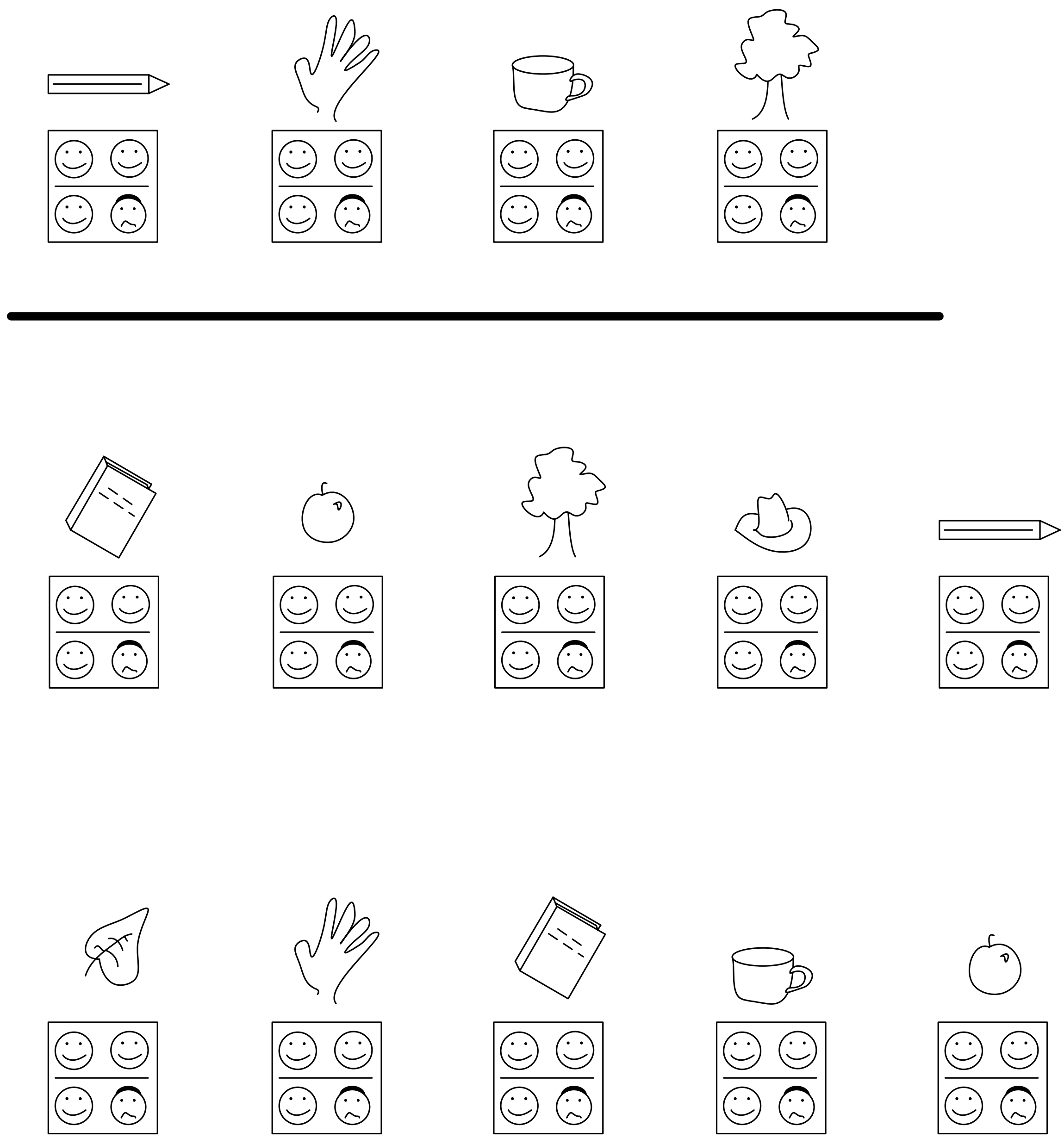
I will be conducting a short research study with fourth graders at the *x* School. The purpose of the study is to explore how fourth graders compose songs. I will use the results of the study to improve my teaching. This study is part of my master's program at Boston University.

Participation is completely voluntary, and your child may stop participating in the study at any time. I will be conducting the study between November 1 and November 24, 2010 at your child's school. Your child's personal information and test responses will be kept confidential. If you would not like your child to participate, please call the school at *xxx-xxx-xxxx* or send a note to your child's teacher.

If you have any questions, you can contact me or my thesis advisor at the phone numbers or email addresses listed below. You may obtain further information about your rights as a research subject by calling the BU CRC IRB Office at 617-358-6115.

Thank you for your participation.

Appendix C – DSD Instrument



Appendix D – Lesson Plan

Objective

The objective of this lesson is to provide students with experience and practice improvising, composing, and performing simple tonal patterns. In addition, students will be given opportunities to practice identifying sameness or difference between two musical stimuli.

Lesson Procedure

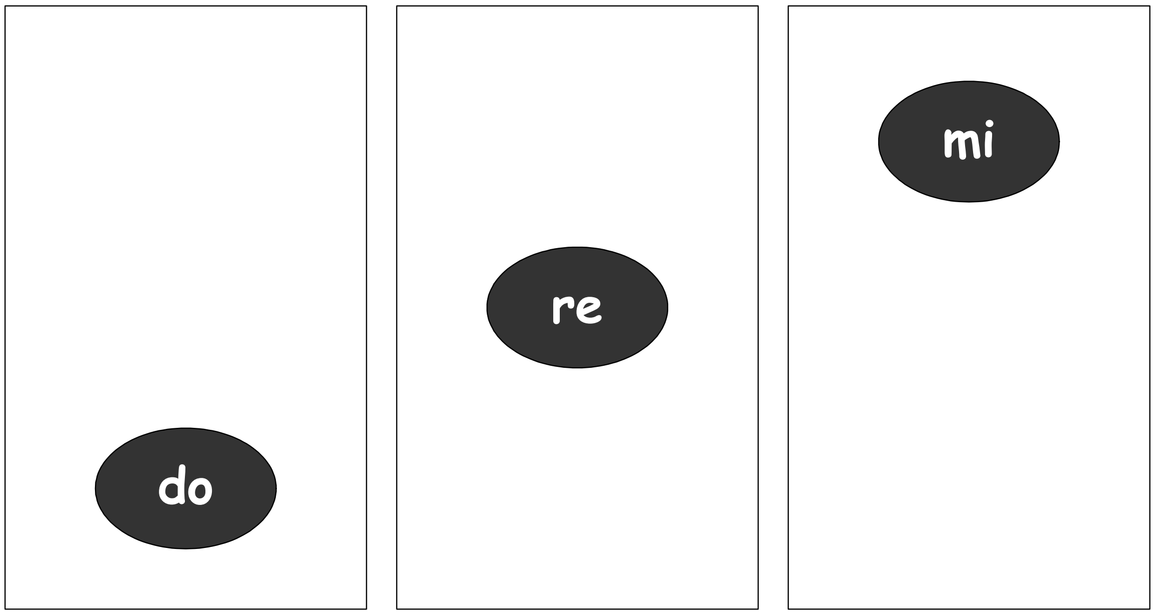
Introduction to lesson (5 minutes): [Teacher-led]. There will be a brief demonstration of proper technique. First, the xylophone and three bars marked with colored dots will be demonstrated. Second, the keyboard and three keys marked with identical colored dots will be demonstrated. Third, there will be a brief echoing activity where the teacher sings a three-note pattern for the students to echo.

Improvisation and echoing in small groups (10 minutes): [Student-led]. Students will be divided into three equal groups: xylophone players, keyboard players, and singers. The xylophone players go first. One at a time, each player in this group performs a three-note pattern. This is immediately echoed by the keyboard players and singers (all together). After all xylophone players have performed and their improvisations have been echoed, the groups rotate until each student has performed with xylophone, keyboard, and voice.

Composition practice (5 minutes): [Researcher-led]. There will be a brief period of instruction where students are taught how to complete the activities they will perform in their individual session. The teacher will show students three large cards, like those used in the individual session. Three volunteers will be chosen to select one pitch each, and these three pitches will be displayed on the board for the class to see. A fourth volunteer will demonstrate tracing the contour, and then the class will sing the pattern in unison.

Discrimination practice (5 minutes): [Researcher-led]. With the pattern composed still on the board, the teacher will ask students to imagine how the contour sounds. Then, the teacher will play the pattern correctly on the piano and ask the class “did that song sound like the contour looks?” Finally, the teacher will perform an incorrect pattern and ask the class the same question. Feedback will be provided as necessary.

Appendix E – Contour Cards



Appendix F – Photos of DSD Session Space

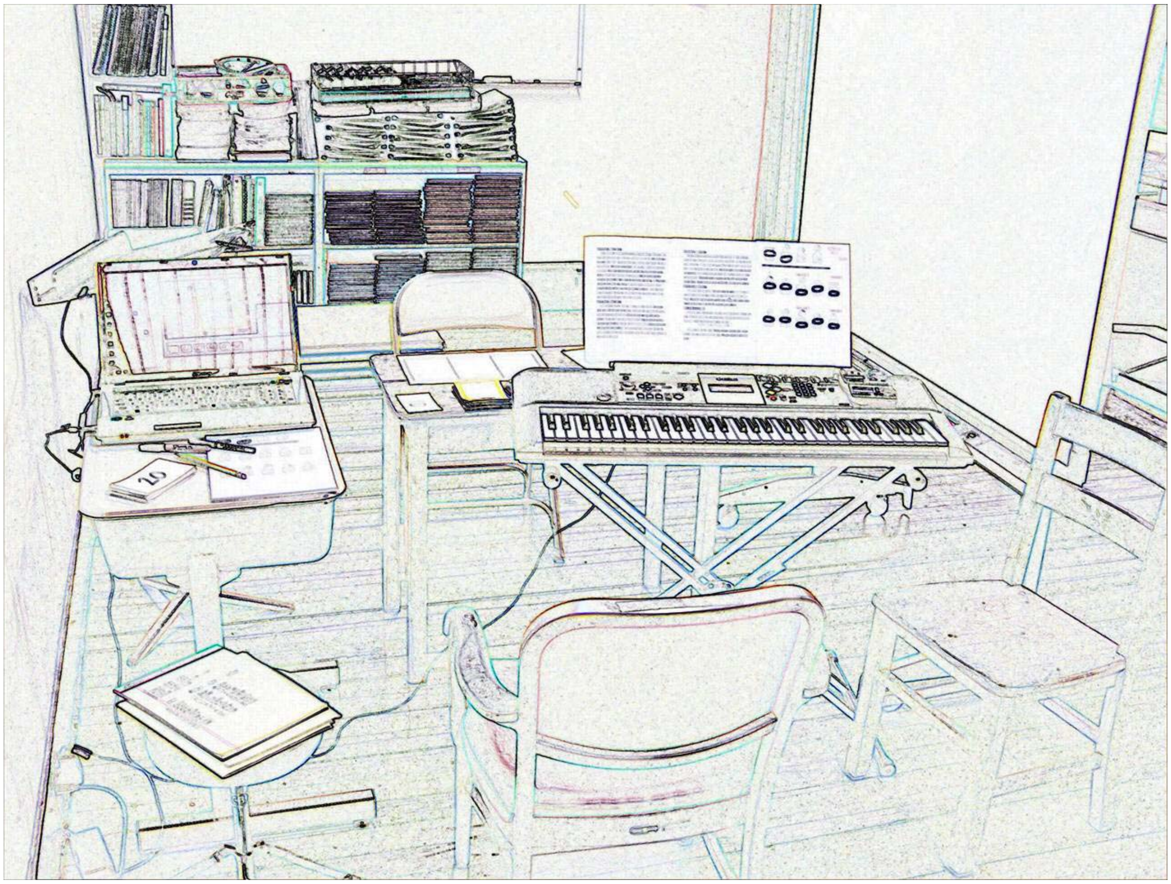


Figure 3 - DSD Session Space, School A, Researcher Perspective[15]

[15] Photographs have been edited to optimize printing.

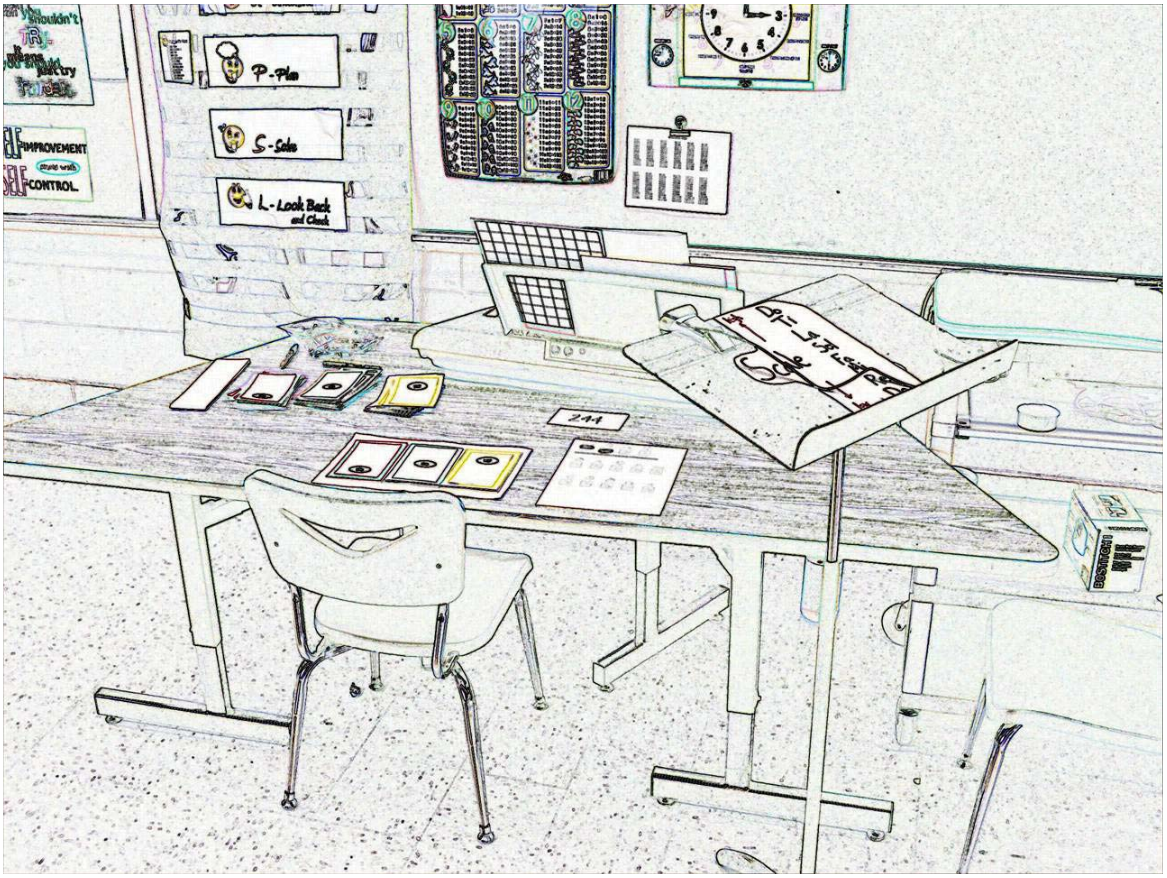


Figure 4 - DSD Session Space, School B, Participant Perspective

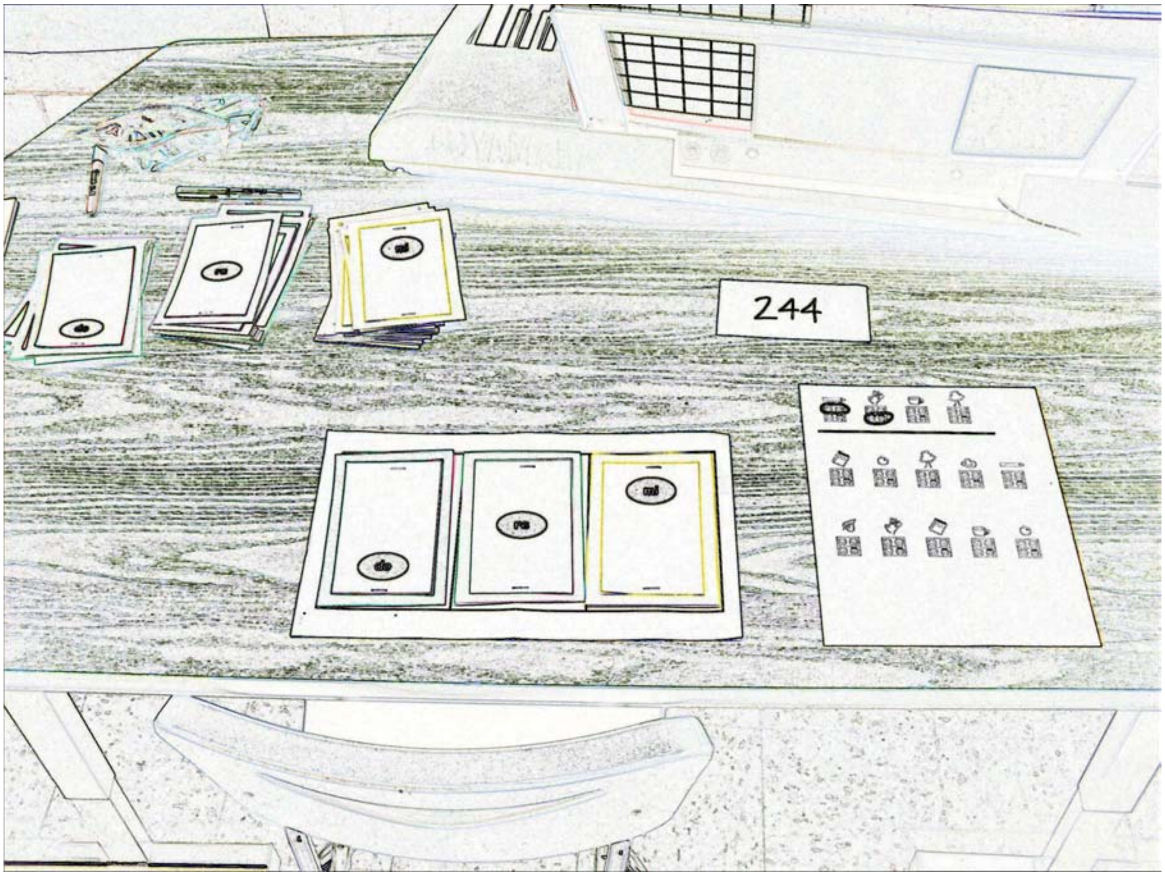


Figure 5 - DSD Session Space, School B, Participant Workspace

Participant workspace.

References

Amabile, T.M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology. 43,* 997-1013.

Auh, M. (1995). *Prediction of musical creativity in composition among selected variables for upper elementary students*. (Doctoral dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 9604632)

Auh, M. & Walker, R. (1991). Compositional strategies and musical creativity when composing with staff notation versus graphic notations among Korean students. *Bulletin of the Council for Research in Music Education. 141*, 2-9.

Azzara, C. D. (1993). Audiation-based improvisation techniques and elementary instrumental students' music. *Journal of Research in Music Education, 41*(4), 328-342. doi:10.2307/3345508

Barrett, J.R. (1990). *Melodic Schemata, forms of representation, and cognitive strategies used by fourth-graders in the recall and reproduction of familiar songs.* (Doctoral dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 9024044)

Cooper, N. (1994) An exploratory study in the measurement of children's pitch discrimination ability. *Psychology of Music 22*(1), 56-62. doi: 10.1177/0305735694221005

Dalby, B., & Gordon, E.E. (2009). *About music learning theory.* Retrieved from http://www.giml.org/

Dammers, R. J. (2007). *Supporting comprehensive musicianship through laptop computer-based composing problems in a middle school band rehearsal.* (Doctoral Dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 3269874)

Davies, J.B. (1979). Memory for melodies and tonal sequences: A theoretical note. *British Journal of Psychology, 70,* 205-210.

Dowling, W.J. (1978). Scale and contour: two components of a theory of memory for melodies. *Psychological Review, 85*(4), 341-354.

Folkestad, G., Hargreaves, D., & Lindstrom, B. (1998). Compositional strategies in computer-based music-making. *British Journal of Music Education, 15*(1), 83-97.

Gaynor, J. (1995). *Music reading comprehension: The effect of aid to chunking and melodic predicting on sight reading performance achievement of secondary school instrumental music students.* (Doctoral Dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 9611506)

Gordon, E. E. (1984). A longitudinal predictive validity study of the Intermediate Measures of Music Audiation. *Bulletin of the Council for Research in Music Education*, *78*(1), 1-23.

Gordon, E. E. (1986). *Intermediate Measures of Music Audiation.* Chicago: GIA.

Gordon, E.E. (1995). *Musical aptitude profile.* Chicago, IL: GIA Publications, Inc.

Guderian, L. V. (2008). *Effects of applied music composition and improvisation assignments on sight-reading ability, learning in music theory and quality in soprano recorder playing.* (Doctoral Dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 3331120)

Hagen, S. (1996). *An examination of the effects of audiation on the compositional processes and products of fifth graders*. Paper presented at the Third International Conference on Technological Directions in Music Education, San Antonio, Texas. Retrieved from http://music.utsa.edu

Han, S. (2004). *Composers' perspectives regarding computer technology's impact on the development of compositional craft*. (Doctoral Dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 3117840)

Hickey, M. M. (1995). *Qualitative and quantitative relationships between children's creative musical thinking processes and products*. (Doctoral Dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 9614754)

Hickey, M. M. (2009). At-risk teens: making sense of life through music composition. In J. L. Kerchner & C. R. Abril (Eds.), *Musical experience in our lives* (pp. 199-215). New York: Rowman & Littlefield Education.

Holahan, J.M. & Saunders, T.C. (1997). Children's discrimination of tonal patterns: Pattern contour, response time, and item difficulty level. *Council for Research in Music Education, 132,* 85-101.

Josuweit, D. (1991). The effects of an audiation-based instrumental music curriculum upon beginning band students' achievement in music creativity. *Dissertation Abstracts International 52*(9), 3213.

Kratus, J. (1989). A time analysis of the compositional processes used by children ages 7 to 11. *Journal of Research in Music Education, 37*(1), 5-20. doi:10.2307/3344949

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. doi: 10.2307/3345496

Massachusetts Department of Elementary & Secondary Education. (1999). *Massachusetts Arts Curriculum Framework.* Retrieved from Massachusetts Department of Education website: http://www.doe.mass.edu

Massachusetts Department of Elementary & Secondary Education (2010). *Student Data*. Retrieved from http://www.doe.mass.edu/profiles

McCord, K. A. (2009). Musical experience in the lives of children with severe disabilities. In J. L. Kerchner & C. R. Abril (Eds.), *Musical experience in our lives* (pp. 127-143). New York: Rowman & Littlefield Education.

McPherson, G. E. (1994). Factors and abilities influencing sightreading skill in music. *Journal of Research in Music Education, 42*(3), 217-231. doi:10.2307/3345700

National Standards for Arts Education. (1994). Reston, VA: MENC.

Prasso, N. M. (1997). *An examination of the effect of writing melodies, using a computer-based song-writing program, on high school students' individual learning of sight-singing skills.* (Doctoral dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 9734090)

Saunders, T.C. & Holahan, J.M. (1993). Computerized response procedure to assess young student reaction times of judgments of sameness and difference among paired tonal patterns. *Council for Research in Music Education, 115,* 31-48.

Stoltzfus, J.L. (2005). *The effects of audiation-based composition on the music* *achievement of elementary wind and percussion students*. (Doctoral dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 3169610)

Strand, K. (2005). *Nurturing young composers: Exploring the relationship between instruction and transfer in 9- to 12-year-old students* (Doctoral dissertation). Retrieved from http://proquest.umi.com/ (UMI No. 3087983)

Strand, K. (2006). Survey of Indiana Music Teachers on Using Composition in the Classroom. *Journal of Research in Music Education, 54*(2), 154-167. doi: 10.1177/002242940605400206

Webster, P.R. (1977) A factor of intellect approach to creative thinking in music. *Dissertation Abstracts International 36*(6), 3136.

Webster, P.R. (1979). Relationship between creative behavior in music and selected variables as measured in high school students. *Journal of Research in Music Education, 27*(4), 227-242.

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

Webster, P.R. (1990) Creativity as creative thinking. *Music Educators Journal, 76*(9), 22-28. doi:10.2307/3401073

Webster, P.R. (1994). Measure of Creative Thinking in Music-II (MCTM-II). Northwestern University, Evanston, IL.

Wilson, S.J., & Wales, R.J. (1995). An exploration of children’s musical compositions. *Journal of Research in Music Education 43*(2), 94-111.

Vita

Travis Michael Ramsey (born 1979) earned a Bachelor of Music in Music Education degree from the University of Southern Maine (USM) in 2003 (magna cum laude). At USM, he studied trombone with Mark Manduca of the Portland Symphony and studied composition with Dr. Joseph Packales and Dr. J. Mark Scearce. His teaching career began in Maine where he served as director of the Freeport (ME) middle school band, director of a Yarmouth (ME) high school jazz combo, and a mentor-composer with the Vermont MIDI Project. In 2004, Travis began teaching elementary general music in a public school district in eastern Massachusetts. Since 2004 he has taught at four elementary schools. In addition to his regular duties as general music teacher, he directs an 80-member elementary chorus, maintains an active, donation-supported band program, serves as music director for the drama club, produces eight to ten annual concerts for the school community, and volunteers his time on the school's Leadership Committee. Travis was a founding member of the district's Community Arts Partnership and has acted as supervising teacher for four outstanding student teachers from the Berklee College of Music.

Travis splits his time between teaching and composing. His works as a professional composer have been commissioned by the Choral Art Society (Portland, ME), the Casco Bay Concert Band (Gorham, ME), the Hartland (VT) Elementary School Band, the Cape Elizabeth (ME) High School Wind Symphony, and the Hanover (NH) High School *Footnotes* select choir. His compositions have been performed by the Wellesley (MA) Symphony Orchestra, the Southern Maine Symphony Orchestra, and several ensembles at the University of Southern Maine, including the Trombone Choir, Chamber Singers, Concert Band, and several outstanding student soloists. Travis is the owner of Kenmore Studios, a small music publishing company. The catalog includes twenty-four compositions for beginning band, as well as his compositions for chamber ensembles, band, chorus, and orchestra.

Travis lives in Malden, Massachusetts with his wife Gisela, their daughter Anna (who is three weeks old as of today), and their two cats Miles and Georgia. He can be reached at the address below.

Travis Ramsey

Kenmore Studios

Malden, MA 02148

tramsey@kenmorestudios.com