The reciprocal relationship between text literacy and music literacy among beginning band students

David Lawson Carroll

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations

Recommended Citation

This Dissertation/Thesis is brought to you for free and open access by the Graduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Graduate Research Theses & Dissertations by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
ABSTRACT

THE RECIPROCAL RELATIONSHIP BETWEEN TEXT LITERACY AND MUSIC LITERACY AMONG BEGINNING BAND STUDENTS

David Lawson Carroll, Ed.D.
Department of Curriculum and Instruction
Northern Illinois University, 2017
Elizabeth Wilkins, Director

This dissertation examines the reciprocal relationship between text literacy and music literacy through an experimental design. Music teachers and English Language Arts (ELA) teachers often address similar components of literacy, including fluency, comprehension, and symbolic interpretation. The theory of cognitive structuralism maintains that through derivative and correlative subsumption, material that is learned in one context is strengthened when applied in a disparate context. Therefore, ELA and music teachers who work in isolation are missing an opportunity to teach parallel literacy concepts for the common advantage of teachers and students. This study seeks to explain how students enrolled in conventional and literacy-enriched band environments perform better than non-band students on text literacy tests. It also asks if conventional band students differ from literacy-enriched band students on text and musical literacy tests.

Fourth-grade students were randomly assigned to one of three groups: a control group \( n = 11 \), a conventionally taught beginning band group \( n = 11 \), and a literacy-enriched beginning band group \( n = 10 \). The experimental treatment included 14 small group band lessons and 14 full band rehearsals. Groups were then compared for textual literacy growth using the NWEA Measures of Academic Performance (MAP) and musical growth using the Watkins Farnum
Performance Scale (WFPS). ANCOVA planned contrasts showed that literacy-enriched students significantly outperformed conventional band students and control students on the MAP literature reading subtest. The research design allowed for the defense of literacy-enriched band instruction as a generalizable cause of higher literacy scores.

Additional comparisons between the control group and the two experimental groups revealed no statistically significant differences between the group means on the overall reading scores or the remaining reading subtest scores. The lack of significance suggested that the statistical model was a poor fit for the data. Furthermore, a small sample size and large unexplained variance contributed to a lack of statistical power. Therefore, the application of the cognitive structuralist theory on the remaining MAP reading tests remained inconclusive.

The conventional and literacy-enriched band groups were also compared against each other on text and music literacy growth. There were no statistically significant differences between the conventional group and literacy-enriched group on the MAP or on the WFPS. This suggested that literacy-enriched instruction in band could benefit a student’s textual literacy skills without compromising musical performance goals. It also suggested that more study is needed to determine the extent to which the explicit instruction of ELA reading skills in band may benefit musical ability.

The overall findings implied that literacy-enriched band instruction caused higher textual literacy scores, and the results may be generalized to similar circumstances. Suggestions for practice included increased collaboration among teachers, enhanced teacher pre-service and in-service opportunities, additional use of student-centered progressive instructional strategies, and the careful reconsideration of eliminating or reducing the availability of band to students. Future research would benefit from the quantitative and qualitative models proposed herein.
THE RECIPROCAL RELATIONSHIP BETWEEN TEXT LITERACY AND MUSIC LITERACY AMONG BEGINNING BAND STUDENTS

BY

DAVID LAWSON CARROLL

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF EDUCATION

DEPARTMENT OF CURRICULUM AND INSTRUCTION

Doctoral Director:
Elizabeth Wilkins
ACKNOWLEDGEMENTS

This work was truly a collaboration among many individuals to whom I am very grateful. First and foremost, I would like to thank my dissertation committee who helped me carve this massive project into something manageable. Dr. Mary Lynn Doherty’s experience and guidance on the experience of beginning musicians and literacy was an invaluable contribution to this text. Dr. David Walker helped me make sense of some rather sophisticated statistical tests, without which I would have been lost. And my chair, Dr. Elizabeth Wilkins, provided the direction, the encouragement, and the resources to turn this dream into a reality. There is simply no one more patient, more thorough, or more inspiring than Dr. Wilkins, and I will be forever indebted to her.

This study was made possible with the financial assistance of the Louise M. Berman Fellows Fund through Kappa Delta Pi, the National Education Association Foundation, and many of my parents, students, and family members who contributed to my research fund online. I am also indebted to the students and teachers at the research site; thank you for taking a risk and agreeing to host my research. I would also like to acknowledge the support and encouragement of my colleagues at Maplebrook and Madison schools, whose thoughtful discussions inspired this research so that more students may have access to quality band and orchestra programs.
DEDICATION

To my lovely bride Carla,

for all of your patience and support…this was truly a team effort.

To my mom and dad, Jane and Wayne,

for teaching me that anything worth doing is worth doing well.

To my students,

for bringing joy to my work every day.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................................................ ix
LIST OF FIGURES ........................................................................................................................................... x
LIST OF APPENDICES ....................................................................................................................................... xi

CHAPTER 1. INTRODUCTION ......................................................................................................................... 1
  Theoretical Framework ................................................................................................................................. 3
  Problem and Purpose Statements ............................................................................................................... 4
  Research Questions ....................................................................................................................................... 5
  Hypotheses ................................................................................................................................................... 6
  Definitions ................................................................................................................................................... 7
  Delimitations .............................................................................................................................................. 9
  Methodology ............................................................................................................................................ 9
  Limitations ............................................................................................................................................... 10
  Significance ............................................................................................................................................ 10
  Organization of Study ................................................................................................................................. 11

CHAPTER 2. REVIEW OF LITERATURE ........................................................................................................ 12
  Introduction and Problem ......................................................................................................................... 12
CHAPTER 4. RESULTS ............................................................................................................. 84

Research Questions .............................................................................................................. 84

Hypotheses ........................................................................................................................... 85

Sample Participants .............................................................................................................. 86

Results for Research Questions 1, 2, and 3 .......................................................................... 94

ANCOVA Assumptions ........................................................................................................... 95

Overall Reading Test Assumptions ....................................................................................... 96

Literature Subtest Assumptions ............................................................................................. 99

Vocabulary Acquisition and Use Subtest Assumptions ......................................................... 99

Informational Text Subtest Assumptions ............................................................................... 100

Omnibus ANOVA Results for the MAP Reading Test and Subtests ................................. 100

Research Question 1: Control Group Versus the Conventional Group ............................ 103

Research Question 2: Control Group Versus the Literacy-Enriched Group ...................... 105

Research Question 3: Conventional Group Versus the Literacy-Enriched Group ............. 108

Results for Research Question 4 ........................................................................................ 109

Summary .................................................................................................................................. 111

CHAPTER 5. DISCUSSION ..................................................................................................... 113

Research Question 1: Interpretation and Discussion .......................................................... 115
Research Question 2: Interpretation and Discussion ......................................................... 120

Literacy Tests: Surface Similarities ........................................................................ 123

Literacy Tests: Rule Out Irrelevancies ................................................................ 125

Literacy Tests: Discriminations ........................................................................... 127

Literacy Tests: Interpolations and Extrapolations ............................................... 129

Literacy Tests: Causal Explanations .................................................................... 129

Research Question 3: Interpretation and Discussion ......................................................... 133

Research Question 4: Interpretation and Discussion ......................................................... 136

Evaluating the Theory: Cognitive Structuralism ............................................................... 139

Limitations ......................................................................................................................... 141

Implications for Practice .................................................................................................... 145

Recommendations for Research ........................................................................................ 148

Conclusion ......................................................................................................................... 152

REFERENCES ........................................................................................................................... 154

APPENDICES ............................................................................................................................ 170
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1: Common Content Area Reading Strategies within Four Texts, 2011-2014</td>
<td>36</td>
</tr>
<tr>
<td>3.1: Frequency of Explicit Strategies Used in the Literacy-Enriched Group</td>
<td>61</td>
</tr>
<tr>
<td>3.2: Scheduling Template for Experimental Groups</td>
<td>62</td>
</tr>
<tr>
<td>4.1: Gender, Ethnicity, and SES as Potential Confounds on Pretest and Posttest Differences</td>
<td>93</td>
</tr>
<tr>
<td>4.2: Classroom Teacher as a Potential Confound on Reading Test Differences</td>
<td>94</td>
</tr>
<tr>
<td>4.3: Overall MAP Pretest and Posttest Reading Scores</td>
<td>101</td>
</tr>
<tr>
<td>4.4: MAP Reading Subtest Pretest and Posttest Scores</td>
<td>102</td>
</tr>
<tr>
<td>4.5: MAP Score Differences Between the Control Group and Group 1 (Conventional Instruction)</td>
<td>104</td>
</tr>
<tr>
<td>4.6: MAP Score Differences Between the Control Group and Group 2 (Literacy-Enriched Instruction)</td>
<td>106</td>
</tr>
<tr>
<td>4.7: MAP Score Differences Between the Group 1 (Conventional Instruction) and Group 2 (Literacy-Enriched Instruction)</td>
<td>108</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure                                                                                      Page
2.1: A holistic conceptualization of language literacy.......................................................... 26
2.2: A holistic conceptualization of music literacy................................................................. 27
3.1: Procedure for experimental treatment, data collection, and equal access to instruction. 59
3.2: Alignment of data collection, data analysis, and research questions............................. 64
3.3: Method for investigating generalized causal inferences by Shadish et al. (2006)............ 81
4.1: Ethnicity of study participants. ........................................................................................ 89
4.2: Free and reduced lunch status of study participants. ....................................................... 90
4.3: Classroom teacher distribution of study participants....................................................... 91
4.4: Scatter plot and regression lines for the overall reading test scores............................. 97
4.5: Scatter plot indicating homoscedasticity between standardized residuals and predicted 98
    values. ................................................................................................................................ 98
4.6: Summary of ANCOVA and independent t test results and their relationship to the four 112
    research questions. ............................................................................................................. 112
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. RECRUITMENT LETTERS</td>
<td>170</td>
</tr>
<tr>
<td>B. PARENT AND STUDENT NOTIFICATION OF RESEARCH STUDY</td>
<td>173</td>
</tr>
<tr>
<td>C. CONSENT AND ASSENT TO PARTICIPATE IN A RESEARCH STUDY</td>
<td>177</td>
</tr>
<tr>
<td>D. CONFIDENTIAL DEMOGRAPHIC SURVEY</td>
<td>182</td>
</tr>
<tr>
<td>E. STUDY BUDGET AND EXPENSES</td>
<td>185</td>
</tr>
<tr>
<td>F. INSTRUCTOR GUIDELINES AND LESSON PLANS TO DIFFERENTIATE BETWEEN</td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL TREATMENTS</td>
<td>187</td>
</tr>
<tr>
<td>G. TRANSCRIPT FROM FOURTH-GRADE COORDINATION MEETING</td>
<td>197</td>
</tr>
<tr>
<td>H. PRE-EXISTING DIFFERENCES BETWEEN GROUPS AND FOURTH-GRADE MEAN</td>
<td></td>
</tr>
<tr>
<td>ON THE MAP READING RIT SCORES</td>
<td>213</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Formal band instruction has enriched the lives of elementary school children since Lowell Mason first introduced music to the public schools in 1838 (Miller, 1984). Since that time, music has occupied a tenuous position as a fixture in public school curricula, even though the Goals 2000: Educate America Act (Bohannon & McDowell, 2010) and the Every Student Succeeds Act (Woodside, 2015) recognize music as a core academic subject. In 2007, instrumental music instruction, which includes band and orchestra, was offered at only 67% of public elementary schools in the United States (National Center for Educational Statistics, n.d.).

The quality and consistency of existing programs differ widely depending on the resources and support of the community. This difference is illustrated by the wide variety of starting grade levels for band instruction. According to Bourne (1993), only 7.7% of school districts begin band instruction in fourth-grade, while 45.4% of districts begin band instruction in fifth grade, 37.7% in sixth grade, and 9.2% in seventh grade. Continuity and common practices are difficult to achieve under such disparate circumstances in curriculum design. Furthermore, the National Core Arts Standards (n.d.) are not grade-specific, and the scopes and sequences they advocate are necessarily vague to provide flexibility for schools without the equipment, resources, or funding to start a formal band program in the fourth-grade.

This inconsistency has become problematic for researchers wishing to identify effective practices in teaching band at the elementary level, particularly among beginner instrumentalists.
Advocates for strengthening music programs have little empirical evidence or generalizations upon which to base their claims, as much of the existing literature investigating the effectiveness of music programs fails to consider the quality or type of music instruction, school resources, or community support as influential factors (Heuser, 2011).

Music is frequently subsumed to other core subjects (Heffner, 2007). With increased emphasis placed on math and literacy understanding because of the Common Core State Standards (CCSS), music is often marginalized or eliminated from the curriculum in favor of additional English/Language Arts (ELA) or mathematics instruction (West, 2012). According to a study performed by Heffner (2007), over 77% of administrators reported knowing students who were required to drop music to add more math or reading.

The marginalization of music instruction in schools has been exacerbated by the inability of music educators to clearly define music’s role within a holistic school curriculum, particularly the relationship between music and ELA. For example, Reimer’s (1970) seminal work on aesthetic education, A Philosophy of Music Education, soundly rejects music as a language, while Hansen, Bernstorf, and Stuber’s (2014) The Music and Literacy Connection embrace the semantical and syntactical constructs that music shares with language.

These formative texts represent the core of two divergent approaches to teaching music and include assumptions that are crucial to this study. For example, if students are to be taught music for its aesthetic benefits, then music literacy, which may or may not lead to aesthetic understanding, becomes a secondary goal. However, if students are to be taught music for its literary benefits, then music performance becomes a secondary goal. Given the nation’s increased focus on inter-disciplinary literacy skills due in part to the emergence of the CCSS, the promotion of music as a language has benefitted the advocacy of the non-musical literary
benefits of music at the expense of the artistic and aesthetic benefits (Gerrity, 2007; Hansen & Milligan, 2012).

Another important distinction between these approaches lies in their treatment of beginners. Most beginner band programs focus on bridging aural readiness with music literacy within the initial phases of instruction (Musco, 2011), although two approaches provide unique paths toward achieving that goal. Conventional instruction, based on aesthetic and perennialist foundations, teaches note-reading through performance. As such, it is mostly teacher-led to help students reach rigorous, time-tested artistic performance standards (Cavitt, 2003; Heuser, 2011; Kratus, 2007). On the other hand, literacy-enriched instruction, based on multi-disciplinary and constructivist foundations, teaches note-reading using a wide variety of student-centered tools and activities (Bazan, 2011). This includes borrowing from the practices and strategies of ELA teachers, such as coding, annotating, summarizing, and predicting (Hall & Robinson, 2012). Despite the emphasis on literacy skills by CCSS advocates, most band directors still utilize the conventional approach in lessons and rehearsals, focusing primarily on the performance of music and not necessarily on the literacy growth of their students (Miksza, 2013).

Theoretical Framework

This study adopted a cognitive structuralist lens as advocated by Bruner (1977, 1987) and Ausubel (2010). This theory proposes that students’ interpretation of a stimulus changes as its structure is applied in multiple contexts (Bruner, 1977). Students may attain a greater understanding of complex ideas by examining those ideas through multiple perspectives and in different applications. For example, students utilizing the same reading strategy in both ELA
and band (such as annotating or coding) may find that strategy more intuitive and effective when practiced systematically in both contexts.

Building upon Vygotsky’s (1934/2012) socio-cultural learning theories, Bruner (1987) and Ausubel (2010) agree that social interaction is key to cognitive development, particularly in language. Band rehearsals rely heavily on socio-cultural interactions (Colson, 2012) and, therefore, provide ideal environments to apply this theory. A cognitive structuralist theory is especially relevant considering musical literacy strategies and textual literacy strategies act as similar functions toward language acquisition. A more detailed examination of the theoretical framework will be given in Chapter 2.

Problem and Purpose Statements

Music literacy is a major component of formal elementary-level band programs (Hansen et al., 2014). However, conventional band instruction, when compared to literacy-enriched instruction, draws from a relatively narrow collection of music literacy strategies. Nevertheless, those literacy strategies implicitly support similar strategies found in ELA classrooms. Possibly as a result of this reciprocity, the literature identifies a fairly strong correlational relationship between music reading and textual reading (Butzlaff, 2000; Winner & Cooper, 2000).

However, the extant literature offers little evidence indicating a causal relationship among music literacy skills and textual literacy skills (Winner & Cooper, 2000). Causal relationships are difficult to establish in the social sciences. Cohen, Manion, and Morrison (2011) described causal influences are more like an inter-related web than a simple linear relationship. But when combined with specific theoretical mechanisms, constructs, and patterns, a causal relationship may be argued. It is unknown, for example, whether textual literacy
achievement is improved because students systematically learn to read music. Improved achievement may alternatively be caused by a combination of family support, socio-economic status, higher intelligence, or gender. Similarly, it is unknown whether explicitly using ELA reading strategies in band causes greater music reading achievement.

This literature gap exists because randomized experimental studies which feature manipulatable independent variables to help mitigate spurious influences are extremely difficult to implement (Butzlaff, 2000; Cohen, Manion, & Morrison, 2011). Withholding band instruction from a large segment of students to establish a control group presents ethical and logistical issues that must be carefully addressed. All students deserve the opportunity to learn a musical instrument. This challenge may be overcome, however, given favorable circumstances, resources, and research design.

Quantitative studies have indicated a positive relationship between music and textual literacy when band directors use primarily conventional teaching approaches. Literacy-enriched approaches would seemingly yield greater growth if the cognitive structuralists are correct (Ausubel, 2010; Bruner, 1977). A controlled experiment may help provide evidence that the combined implementation of literacy strategies in ELA and music causes observable growth in both areas. Therefore, the purpose of this study was to examine the reciprocal effects of music literacy instruction and textual literacy instruction among elementary band students in conventional and literacy-enriched contexts.

**Research Questions**

The following research questions concern textual literacy achievement:
1. How does conventional band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?

2. How does literacy-enriched band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?

3. Does literacy-enriched band instruction yield a different effect from conventional band instruction on textual literacy scores among elementary students?

The following research question concerns music literacy achievement:

4. Does literacy-enriched band instruction yield a different effect from conventional band instruction on musical literacy scores among elementary students?

Hypotheses

Hypothesis 1: Students participating in a conventional band environment will exhibit greater textual literacy growth than students who do not participate in band when controlling for pretest scores. In other words, \( H_1: (\mu_1 > \mu_0) \) where 0 is the control group (non-band) and 1 is the conventional band group. The null hypothesis states that no difference exists between the groups or the control group outscores the conventional band group, or \( H_0: (\mu_1 \leq \mu_0) \).

Hypothesis 2: Students who participate in a literacy-enriched band environment will exhibit greater textual literacy growth than students who do not participate in band when controlling for pretest scores. This is represented with \( H_2: (\mu_2 > \mu_0) \) where 0 is the control group (non-band) and 2 is the literacy-enriched band group. The null hypothesis states that no
difference exists between the groups or the control group outscores the literacy-enriched band group, or H0: (µ2 ≤ µ0).

Hypothesis 3: Students in the conventional band group will exhibit different textual literacy scores than students in the literacy-enriched band group when controlling for pretest scores. The null hypothesis suggests that no difference exists between the group means. Otherwise stated, H3: (µ1 ≠ µ2) where 1 is the conventional band group and 2 is the literacy-enriched band group, while the null hypothesis is similarly coded H0: (µ1 = µ2).

Hypothesis 4: Students in the conventional band group will exhibit different musical literacy scores than students in the literacy-enriched band group. The null hypothesis suggests that no difference exists between the group means. Otherwise stated, H4: (µ1 ≠ µ2) where 1 is the conventional band group and 2 is the literacy-enriched band group, while the null hypothesis is coded H0: (µ1 = µ2).

Definitions

For the purposes of this study, the following terms are defined:

**Achievement vs. Growth** – As defined by Cronin, Kingsbury, McCall, and Bowe (2005), achievement refers to the empirical evidence of improved knowledge and skills as measured by a test or during systematic observations (i.e., formative and summative assessments) at a single point in time. Growth refers to the difference in scores or observations among several points in time.

- **Textual Literacy Growth** is defined as an increase in reading scores on the Northwest Evaluation Association’s Measures for Academic Performance.
Musical Literacy Growth is defined as an increase in music reading scores on the Watkins-Farnum Performance Scale.

Literacy – Literacy refers to the ability to “articulate the practices and foundations that are valued in a given discipline or profession” (Hansen et al., 2014, p. 3). This definition includes the patterns of reasoning and information processing associated with the interpretation of semiotics and phonemics, as well as means of communicating ideas, such as reading, writing, listening, and speaking (Harris & Hodges, 1995; McIntire, 2007).

Strategy – Strategy refers to specific tasks, activities, and approaches taken during the course of a single lesson or unit to “maximize the possibility of enhancing student achievement” (Marzano, Pickering, & Pollock, 2004, p. 3). This may include peer-evaluation, a vocabulary recognition game, or a repetitive task for learning a technically difficult passage on an instrument. Strategies are manifestations of the orientations that guide their purpose and outcome.

Conventional Band Group – The conventional band group is the experimental group utilizing conventional band instructional techniques (see Appendix F). Typical techniques include trial and error practicing, teacher/peer modeling, verbalizing note names, verbalizing fingerings, audiation, and repetitive drill (Conway, 2003; Henninger, Flowers, & Councill, 2006). This group is labeled Group 1 on the statistical analysis documentation.

Literacy-Enriched Band Group – The literacy-enriched band group is the experimental group that will use conventional instructional techniques blended with literacy-enriched strategies borrowed from typical English/Language Arts (ELA) applications (see Appendix F). These techniques include using codes, annotations, visualizations, front-loading, think-alouds, and read-alouds (Daniels & Zemelman, 2014). This group is labeled Group 2 on the statistical analysis documentation.
Delimitations

The target population consisted of non-special education fourth-grade students ages nine and ten, since this age group typically comprises beginning participants of instrumental music classes at the youngest level. Additionally, this study investigated textual literacy growth in the areas of literature comprehension, informational text comprehension, and vocabulary, as opposed to written, mathematical, oral-linguistic, affective, or other non-musical outcomes resulting from music instruction. Finally, this study only investigated the effect of band instruction, not general music, orchestral music, choral music, music theory, or emerging ensemble instruction such as group guitar or keyboard classes.

Methodology

This study utilized a quantitative experimental research design. The participants consisted of beginning band students chosen from an urban elementary school using a randomized sampling design. Data were collected by dividing students into a control group, a conventionally taught band group, and a literacy-enriched band group. The groups were compared on text reading growth and music reading growth using standardized tests.

Students completed a literacy-based pretest and posttest consistent with a single factor multiple-treatment design. Data from the literacy test were analyzed with an analysis of covariance (ANCOVA). The dependent variable was the posttest reading score, and the factor was the type of instruction received. The posttest scores were controlled using the pretest scores as a covariate. Means differences among posttest scores were adjusted by the covariate then investigated to determine if the group means differed by treatment type on literacy outcomes.
The instrumental performance-based test was conducted for the two band groups using a one-shot case study design. Data from the instrumental performance test were analyzed with an independent samples \( t \) test. The treatment type (conventional or literacy-enriched) was the independent variable while the music performance test score was the dependent variable. Means differences among performance scores were investigated to determine if the group means differed by treatment type on musical outcomes.

Limitations

This study was limited by the generalizability of the findings due to the research design and quantitative methodology, which does not easily account for complex social behaviors (Mertens, 2015). The study was also limited by a low sample size due to the diminished availability of sites able to administer the prescribed program. Lastly, the study was limited by available funding, as the cost of stipends, facilities, and instruments prevented the implementation of this study among a wider variety schools and districts.

Significance

This study provided a necessary base of experimental research focused on reducing extraneous variables from influencing the study’s outcome. As of yet, a fair number of quantitative studies indicate that band may positively influence reading scores, but they have been unable to control latent variables or provide randomized samples that would have inferred causal relationships. Therefore, their conclusions cannot generalize causation with confidence. This study may be the first causal experimental design involving band participants in the literature.
Curriculum directors and policy makers benefit from this study, as it investigated whether band participation can improve textual literacy growth. Band classes are often eliminated or marginalized since their academic impact is difficult to measure (West, 2012). The current study suggests that reducing the opportunity to apply literacy strategies in a non-ELA context (music) may not benefit students as much as retaining them in their band program.

This study also benefits band instructors by indicating that applying textual literacy strategies to music instruction will not hinder band students’ progress to become better music readers and music performers. Also, band instructors and ELA instructors could mutually benefit from increased collaboration when teaching parallel concepts. Furthermore, students will be able to apply literacy strategies in greater depth and in multiple contexts, leading to better long-term understanding in both music and ELA.

Organization of Study

This study is offered in five chapters. Chapter 1 presents an introduction and overview of the problem. This includes the research questions and the significance of the study. Chapter 2 includes a review of the literature relevant to music teaching and literacy. Chapter 3 describes the methodology used for the study. Chapter 4 presents the quantitative data that were collected throughout the study. Lastly, Chapter 5 conducts a discussion of the findings, the implications, and recommendations for future research.
CHAPTER 2

REVIEW OF LITERATURE

Introduction and Problem

The relationship between literacy instruction and music instruction directly reflects America’s ever-changing epistemological, axiological, and political views. A product of the democratic process, education reform initiatives attempt to resolve tensions caused by the perceived faults of equity, access, and economic opportunity with each new election (Tyack & Cuban, 1995). As a result, the delicate balance between the resources allocated for literacy instruction and for competing content areas is often disrupted (Darling-Hammond, 2010).

Supreme Court rulings and federal education legislation have attempted to rectify those faults of equity, access, and opportunity through court decisions such as Brown v. Board of Education (1954), public acts such as The Elementary and Secondary Education Act of 1965 (ESEA) and the Act’s substantial reauthorization in 1994 (Improving America’s Schools Act, Tyack & Cuban, 1995), the 2002 No Child Left Behind Act (Darling-Hammond, 2010), Obama’s signature education initiative Race to the Top (RTTT, Branscome, 2012), and the 2015 Every Student Succeeds Act (Woodside, 2015). The cumulative effect of these actions has left many content areas vulnerable, as mutable priorities influence educational reforms (Apple, 2000). The enduring consideration of music as a core academic subject largely depends on the reculturalization of society about music’s important artistic and academic benefits.
Overview

This literature review will begin by exploring the political and axiological contexts surrounding the state of contemporary music education, including a description of the traditional and progressive orientations for instruction. Next will follow the theoretical implications of cognitive structuralism as advocated by Jerome Bruner (1977, 1987, 1997) and David Ausubel (2010). The literature review will then continue with an exploration of literacy in various contexts, including the proposal of a holistic conceptual model that accommodates diverse perspectives of literacy understanding. Finally, the literature review will explore related empirical studies that have previously investigated the relationship between music and academic achievement among elementary and middle school students.

Political and Axiological Context of Music Education

The history of music education in the United States has been marked with conflicting axiological beliefs about how and why music is important in society. The terms aesthetic, intrinsic, or essentialist philosophies are used interchangeably in the literature while retaining a practically similar meaning. They state that music must be taught and consumed for its own beauty and personal fulfillment and are championed by notable pedagogues such as Elliot (1995), Leonhard and House (1959), and Reimer (1970). The terms pragmatic, extrinsic, or instrumental philosophies are also used interchangeably in the literature. They state that music is interconnected through other content areas and the non-musical benefits of music-learning are as important or nearly as important as the musical benefits. The instrumental philosophies are
promoted by the National Association for Music Education (NAfME) in its Broader Minded campaign (n.d.) and Advocacy Position Statement (2015).

A historical overview of music education reflects a continuous pattern of actions and reactions balancing the essentialist and instrumentalist justifications for inclusion of music in the school curriculum (Huber, 2009; Johanningmeier, 2010; Miller, 1984; Sturm, 1998). Contemporary federal education policy has forced similar adjustments by the music education community. In many instances, music programs are reduced or eliminated, music teachers are required to teach non-music subjects, academically low-achieving students are precluded from participating in music, and the overall time allotted for music is reduced (West, 2012). The reaction to federal reform initiatives by the music education community has been studied by Branscome (2012), who asserted that music advocates are continuously adapting their message to fit the dominant political power, and the following literature supports those findings.

NCLB (2002) has been particularly impactful since it required punitive action against schools that failed to meet adequate yearly progress regardless of the school’s socio-economic health, family composition, community resources, or teacher experience. Several quantitative studies used surveys to examine the attitudes of administrators concerning their schools’ music programs after NCLB, and the findings do not fare well for music programs. For example, Gerrity (2007) investigated the attitudes of 179 principals in Ohio and found that excellent and effective school principals viewed their music programs more favorably than academic watch and academic emergency school principals. He also suggested that 43% of Ohio principals reported weaker music programs because of lost instructional time due to NCLB-necessitated responses.
That same year, Heffner (2007) investigated the impact of NCLB on music programs by surveying 214 music supervisors around the United States. He found that 77% of surveyed schools reported that students were required to drop out of music to take a reading or math course, 52% reported that teachers stopped music instruction to teach test preparation skills, and 26% of music instructors were required to teach reading classes. A third survey study performed by Abril and Gault (2008) found that low socio-economic schools offered far fewer music courses. Additionally, they found that NCLB and standardized testing are considered the biggest obstacles to a strong music program, according to 541 principals across the United States. All three survey studies conclude that many people correlate NCLB legislation with a significant marginalization of music.

Two studies used qualitative methods to examine the effects of NCLB on music programs through phenomenological research methods. Newberg-Long (2010) investigated three teachers’ perceptions of the impact of NCLB and found that lack of instructional time and resources were major stressors, especially in science and music, PE, and art. She also determined that teachers felt that scripted, inflexible curricula were not meeting students’ needs. They advocated for integrated units built around student interests. Her findings support the need for music instruction to provide the holistic, student interest-based instruction lacking in scripted curricula.

Spohn (2008) performed a mixed-methods, sequential quantitative-first study on a Title I school district in Ohio. She found that because of NCLB, educators have improved their efforts to test and retest students until they showed proficiency in math and English/Language Arts (ELA). However, this practice took time away from other content areas. She contends that a
better system would allow local control over assessment practices, even though this would mean
de-standardizing the goals and content of math and ELA Common Core State Standards.

The nature of music assessment provides one reason music is marginalized under NCLB. Some consider music to be an aesthetic art form which is difficult to assess quantitatively (Eisner, 2001; Heckman, 2011). Additionally, music learning conditions are vastly inconsistent between school districts (Parsad & Spiegelman, 2012), and aesthetic outcomes are not clearly associated with economic benefits (Nussbaum, 2010). Not surprisingly, the empirical evidence relating music education to ELA education is too inchoate to overcome these deficits, so the nature of music’s relationship to literacy still remains undefined, necessitating more research in the field.

President Obama chose to eliminate the punitive actions of NCLB by enacting his Race to the Top (RTTT) initiative. RTTT incentivized states to continue using standardized tests and accountability measures in return for federal funding incentives. As a result, the core issues facing music education, i.e., time allocations, high-stakes testing, and instructional inconsistencies, were not remedied by RTTT, and in some cases worsened (West, 2012). The Every Student Succeeds Act (ESSA) reduced the requirements for testing and empowered states and teachers to play a greater role in policy decisions (Woodside, 2015). However, the implementation of ESSA in the Trump era is largely uncertain under the leadership of pro-school choice advocate and Department of Education Secretary Betsy DeVos (Camera, 2017).

In summary, both quantitative and qualitative data suggest that music has been marginalized within the era of accountability and standardized testing characterized by the adoption of NCLB (2002). The challenge lies with the inability of music advocates to adapt music (as a compulsory course or as an elective course) into the continuously fluctuating
axiological context of legislated academics. Alternatively, perhaps the challenge lies in convincing the public that music is too important to be subjected to the whims of political zealotry.

**Traditional versus Progressive Orientations**

As essentialist and instrumentalist justifications for music are reflected in a parallel disagreement between traditional and progressive instructional orientations, essentialists trend toward a traditional orientation, while instrumentalists trend toward a progressive orientation (Miksza, 2013). Traditional band instruction does not easily lend itself to the adaptation of literacy-based orientation. Traditional instruction is defined as teacher-centered and performance-based. Its opposite is progressive instruction, which is student-centered and uses multiple modes of formative and summative assessment. Progressive instruction creates an environment necessary for cross-curricular, multi-sensory, constructivist experiences. For the purposes of this study, literacy-enriched instruction will be considered a sub-set of progressive instruction which considers the holistic importance of both contextual and communicative functions of literacy. Music educators should clearly understand the progressivist conditions that are ideal for advancing English/Language Arts (ELA) concepts in the music classroom through research-based practices.

The literature identifies an expressed interest among educators to systematically examine instrumental music rehearsal orientations and strategies to make instruction more efficient and impactful. Nearly all of the literature regarding band rehearsal effectiveness systematically investigates the individual components of the traditional rehearsal, such as time usage (Goolsby, 1996, 1997), concept prioritization (Goolsby, 1999; Silvey, 2014), error identification (Cavitt,
2003), and assessment (McCreary, 2001; Silvey, 2014). Each study follows the assumption that the performance at hand is the ultimate goal and that more time actually playing an instrument inevitably translates to greater understanding, i.e., a traditional approach to full band rehearsals.

But these strengths may also be seen as a weakness through a progressive lens. There is some question as to how the traditionalists would accommodate learners who are not developmentally ready to perform at the same level as their peers (see Kratus, 2007). Similarly, when considering that band is an elective course, how can students be motivated to accept or even seek constructive criticism without alienating them or causing them to quit altogether (see Bazan, 2011)? Finally, can directors better promote long-term understanding by teaching conceptually rather than teaching to the specific product (performance) at hand (see Misenhelter, 2000)? Progressive approaches help solve these issues through a flexible, meaningful student-centered focus of instruction.

Progressive instruction is not without its critics. Despite the advantages of progressive instruction, 80% of music teachers and 93% students still see traditional performance tests and written tests as the most appropriate for school ensembles, with only 5% of respondents indicating that portfolios, a progressive idea, would be appropriate for school ensembles (McCreary, 2001). Alternative means of assessing students (e.g., projects, compositions, arrangements, and videos) are still not considered valid assessments by most band instructors (Miksza, 2013).

A small but growing body of literature advocates for the progressive instruction of vocabulary, metacognition, coding, annotation, and other textual literacy-based strategies in music classes. The few include Merrill’s (2002) conceptual article “Successful Singing for All in the Elementary Grades,” which explores teaching music vocabulary seamlessly within an
authentic musical context. Similarly, Leonhardt (2011) suggested that music teachers use a word wall for learning music vocabulary, and Benton (2013) discussed how to better promote metacognition in school music classes. Hansen and Bernstorf have published music and text literacy articles separately (e.g., Bernstorf, 2013; Bernstorf, 2016; Hansen & Milligan, 2012) but their findings are summarized well in The Music and Literacy Connection (Hansen et al., 2014), which provides a comprehensive exploration of ELA progressive orientations in the context of music reading. Beyond these examples, little literature is available.

Theoretical Implications of Cognitive Structuralism

Cognitive structuralism emerged as a response to the seemingly untenable and unparsimonious views of the neobehaviorists (Ausubel, 2010). Neobehaviorist theorists such as Tolman, Hull, and Skinner advocated a logical positivism that sought truth through observable behaviors and responses to controlled environments. On the contrary, cognitivists purported that meaning is not a response, but a “conscious experience that emerges when potentially meaningful signs, symbols, concepts, or propositions are related and incorporated within an individual’s cognitive structure on a non-arbitrary and nonverbatim basis” (Ausubel, 2010, p. 40). The cognitivists valued thinking, problem-solving, and concepts over stimuli, responses, and classical and operant conditioning.

Ausubel (2010) and Bruner (1977, 1987, 1997) were particularly influential cognitivists. They suggested that meaningful reception learning is the ideal learning process for an institutionalized (school-based) education. This type of learning features non-arbitrary content that learners can relate to existing anchoring concepts already present in their cognitive structure. As new material is related to existing structures, concepts are reconciled and assimilated into
different, more inclusive structures. In this regard, learning may be defined as the creation and strengthening of cognitive structures that reinforce the core ideas of any given discipline (Bruner, 1977). In other words, “knowledge one has acquired without sufficient structure to tie it together is likely to be knowledge that is forgotten” (p. 31).

This structure is built through the process of correlative and derivative subsumption. Specifically, correlative subsumption extends single concepts that are already present within a learner’s cognitive structure, while derivative subsumption creates new material through the interaction and linkage of multiple concepts in the learner’s cognitive structure (Ausubel, 2010). Both subsumption types are utilized in the learning process, and both reinforce long-term understanding.

Correlative and derivative subsumption may be triggered two ways, through reception learning or discovery learning. Reception learning features a learning environment in which students interpret more denotative, or literal, types of knowledge in a formalized school setting, and is more practical and efficient in the typical classroom than discovery learning. Discovery learning, on the other hand, is more common and appropriate for non-school or non-structured environments, typical of connotative or non-literal types of knowledge (Ausubel, 2010).

Reception learning may be divided into meaningful learning (non-arbitrary) or rote learning (arbitrary). Rote learning does not result in the acquisition of meaning or the long-term strengthening of cognitive structures because its illogical or random nature does not allow new connections to existing cognitive structures. Tasks such as memorizing arbitrary vocabulary words, unrelated flash cards, and/or decontextualized historical facts or geographic locations are examples of rote learning that are not meaningful unless they are anchored to pre-existing ideas (Ausubel, 2010).
Reception learning also contrasts with (but is frequently confused with) expository verbal instruction, or lectures, in which students sit passively and receive knowledge without processing, interpreting, integrating, or applying it. Meaningful reception learning, like discovery learning, is an active process in which students relate, reconcile, and link new material to existing structures. Ausubel (2010) was careful to note, however, that empirical evidence may not be present to indicate student understanding. In fact, the growth of cognitive structures may consist of purely mental processes that have no observable evidence and may, therefore, be difficult to assess using traditional methods.

**Application of Theory in Related Studies**

In their music and text literacy meta-analysis, Winner and Cooper (2000) acknowledged the strength of the cognitive structuralist theory as a mechanism for describing the reciprocal relationship between music reading and textual reading by stating, “The most direct link from learning in the arts to learning in other disciplines is a link in cognitive structure” (p. 12). Standley (2008) adds, “Rather than replicating the techniques of the classroom teacher or reading specialist, the music educator could add a more valuable dimension to the reading curriculum by being trained in the incorporation of reading skills into musical activities” (section 5, para. 11).

Gromko’s (2005) study relating music instruction with phonemic awareness features a theoretical framework based on Bruner’s (1977) theory of cognitive development. Her conclusion is particularly poignant:

A rigorous test of the hypothesis that children may benefit from a second symbol system (e.g., language) from the exercise of the first one (e.g., music) will require a particular kind of music instruction, one that involves active music-making and the association of sound with developmentally appropriate symbols. The implication for schools is that
music instruction, while valuable for liberating the artistic and musical potential of every child, may significantly enhance children’s language literacy as well. (p. 207)

**Application of Theory to Current Study**

The act of interpreting symbols and concepts constitutes the core of language development, whether the language is textual or musical (Hansen et al., 2014). The field of language arts and content-area literacy has enjoyed a sustained history of the development and advocacy of specific strategies designed to enhance understanding in the realm of textual language. However, music literacy does not have the same quality and depth of research as does text literacy (Hansen et al., 2014). Music educators have generally missed an opportunity to strengthen and build on cognitive structures related to language development.

Cognitive structuralism explains theoretically why quasi-experimental, causal-comparative, correlational, and qualitative studies have shown positive relationships between music participation and standardized reading test scores. Since confounding factors or independent variables were not controlled in those studies, the true cause of the positive relationship could not be determined. However, an experimental environment in which confounding factors are randomly distributed across all groups may help clarify whether music instruction causes higher reading scores and whether reading instruction causes greater musical ability by isolating the treatment as a mediating variable in the causal equation.

**Exploration of Music Literacy**

An in-depth investigation into the relationship among music literacy and language literacy must entail an analysis of what is included and what is excluded in an operational
definition of music literacy. Research confirms the complexity of the construct, as the concept of music literacy is interpreted differently among disparate social/cultural traditions and age/skill levels (Mills & McPherson, 2015). Two foundational concepts will be defined to begin this investigation: first, music and language are related, and second, both music literacy and language literacy must be viewed in their molar, or all-inclusive, form.

These concepts will be characterized in a conceptual model of language literacy and music literacy that includes both discipline-specific background knowledge and communication skills. The presentation of the model will be followed by evidence showing how prevailing research supports it in both Western and non-Western cultural traditions. Non-Western traditions include world music developed and communicated outside of the conventional European notation system (Nettl, 2015). Finally, research will show how the model reflects differentiated views by educators in Western cultures considering student age and skill levels.

**Music Literacy and Language Literacy as Related Processes**

There are few sources of disagreement among music educators greater than the debate over the definition of music literacy, principally because no one standard definition fits all ages, time periods, cultures, or philosophies (Hansen et al., 2014; Mills & McPherson, 2015). Instead, music literacy is adapted and interpreted subjectively to fit the conceptual interpretation and experience of the educator. As a result, American music educators often artificially limit their concept of literacy as a reflection of their pedagogical training, missing the opportunity to adopt more pluralistic and student-centered approaches to literacy (Tremblay-Beaton, 2015).

Evidence increasingly shows that music literacy and textual literacy are overlapping and intersecting processes. In other words, language and music are inherently related. However, this
concept was refuted by Leonhard and House (1959), who advocated that music should be taught for its own beauty and personal fulfillment rather than being limited to the rules of language. Agreeing, Reimer (1970) famously stated, “Music…is not in any sense a language. It is neither a non-verbal language (such as numbers or musical notation or dots or dashes, etc.), nor an indefinite language nor a language of the emotions” (p. 32). Reimer claimed that arts education was necessarily non-conceptual, focused instead on subjective aesthetic perception untenable by the confines of language (Allsup & Lewis, 2015). Manins (2001) added that music does not convey clear meaning as does language. The interpretation of music is too relativistic and abstract to be used for communication. He argued that music is aesthetic at its core, but music notation is an insufficient medium to communicate emotion or beauty to a reader.

However, sufficient neurological evidence shows that music processing and language processing share common physiological artifacts. The mechanisms that regulate text processing are the same as the mechanisms that regulate music processing, especially regarding the discrimination and perception of music and text sounds (Asaridou & McQueen, 2013). Continued development and practice of music processing skills have been shown to strengthen the arcuate fasciculus area of the brain, which corresponds to speech and language processing (Halwani, Loui, Rüber, & Schlaug, 2011). Aural perception has also been strongly correlated to auditory discrimination of musicians, which in turn helps strengthen long-term memory (Gromko, Hansen, Tortora, Boccia, & Higgins, 2009). Kraus’s Auditory Neuroscience Lab found that playing a musical instrument aids neural processing because powerful cognitive processes are combined with the emotional reward of playing an instrument (Strait & Kraus, 2014). Thus, the physiological evidence linking music and language provides a basis for reconsidering Reimer’s (1970) aesthetic philosophy.
Rejections of the anti-language aesthetic justifications of music were also made on philosophical grounds. Elliott (1995) criticized Reimer as myopic because he neglected to address musical context; non-Western musical traditions; the sociological, political, and cultural meanings of music; and the epistemological significance of performing and making music (McCarthy & Goble, 2005). Allsup and Lewis (2015) called Reimer’s position unstable and contradictory because his rejection of formalism was tempered with an acknowledgment that meaning is derived from an inherent aesthetic structure.

Likewise, when considering the functional relationship among language literacy and music literacy, the literature reveals many similarities. Both literacies require higher-order processing and abstraction (Cantwell & Millard, 1994). They also require an awareness of pitch recognition and perceptual attunement, which combine context with frequency to infer meaning (Asaridou & McQueen, 2013). Both require a continuum of learning, featuring aural awareness, focused attention, and purposeful repetition (Bernstorf, 2013). Western cultures often introduce coding systems at an early age, reinforcing the sound to symbol to sound cycle (Bernstorf, 2013; Mills & McPherson, 2015). Coding systems include the interpretation of icons and symbols into meaningful information. Language literacy and music literacy share these features.

Furthermore, the aesthetic philosophies of Leonard and House (1959), Reimer (1970), and Manins (2001) lack pragmatism in American elementary and middle schools. Large performance music class sizes, coupled with socio-political pressures to improve reading skills in all content areas, make expressive and aesthetic learning difficult and costly, especially among young children (Barlar, 2010). Aesthetic reactions are often based on subjective individualistic experiences that are difficult to foment consistently in a large group environment. Performance music classes also tend to be based on professional models that mostly value performance and
reading, rarely valuing creativity and contextual background knowledge (Heuser, 2011; Tremblay-Beaton, 2015). Despite the ultimate goal of aesthetic appreciation, young students in the school setting learn more efficiently and effectively by using a common means of communication. In this sense, music does work much like a language.

Conceptual Language and Music Literacy Model

A second foundational concept in the investigation of music literacy is that music literacy must be viewed in its most molar sense, consisting of many parts (Hansen et al., 2014). Most literature regarding language literacy fits into a multi-dimensional conceptual model of literacy (see Figure 2.1). The model illustrates the molar nature of many contemporary definitions of

![Language Literacy Prototype](image)

Figure 2.1. A holistic conceptualization of language literacy.
literacy, particularly since the development of the Common Core State Standards. This model is applicable for both ELA and non-English/Language Arts discipline areas (Urquhart & Frazee, 2012). Relatively narrow definitions that limit literacy to reading and writing are not wrong; they are simply incomplete, representing only one or more components within the greater conceptual model. Literacy encompasses both discipline-specific contextual background knowledge (shown on the left) and communications skills (shown on the right). Music literacy is a derivative of the language literacy prototype and contains nearly identical organizational principles and terminologies (see Figure 2.2).

![Figure 2.2. A holistic conceptualization of music literacy.](image)

Affective responses are missing from this model, even though the literature regards emotion and attitude as an important product of music study (Schubert & McPherson, 2015). Certainly, emotional responses provide enjoyment and motivation for music learners. By many
accounts, music’s affective response may be the most important element of a students’ musical experience (Behne, 1997; Eisner, 2001). However, emotional responses are difficult to observe and measure, as they heavily rely on personal context and experience. Their intensities may vary among people, or within the same person over time. Because of this uncertainty and unreliability, affective experiences are difficult to include in a model that seeks to create a consistent and reliable framework for defining literacy. As a result, the model follows Jorgensen’s (1981) recommendation to separate affective and emotional components from the empirical and objective components of knowledge, particularly in music.

**Deconstructing the Model**

The language literacy and music literacy models were developed based on the combined theories of several experts, which will be discussed throughout the deconstruction of the model. Language literacy encompasses a wide-range of skills and knowledge beyond reading and writing. Freire (1987) advocates holistic approaches to literacy that recognize cultural and social contexts as critical components to learning. Darling-Hammond (2010) defines language literacy as the students’ ability to transfer what they have learned to new problems. Learning transfer is made possible by understanding the fundamental concepts of the content (Bruner, 1977) and the mechanics of communicating those concepts (Shanahan & Shanahan, 2012). Shanahan and Shanahan (2012) have written extensively about the inclusion of discipline-specific skills and knowledge within literacy development. Additionally, Vacca, Vacca, and Mraz (2011) describe literacy as “a synthesis of language, thinking, and contextual practices through which people make and communicate meaning” (p. 7).
Similarly, music literacy also encompasses a wide range of skills and knowledge beyond performing and composing. Mills and McPherson (2015) argue that music literacy “occurs as a result of children having developed their capacity to make music, reflect on the music in which they are engaged, express their views on music which they play, hear, or create, speak about and listen to music to form judgements, and read, write, comprehend, and interpret staff notation” (p. 155). Tremblay-Beaton (2015) described the multi-dimensional view of music literacy succinctly:

Just as music making involves more than decoding signs and symbols, being literate involves much more than simply knowing how to operate the language system. It also includes the cultural and critical facets of knowledge integral to being literate. Freire’s view of literacy is one that is not exhausted merely by decoding the written word or written language, but rather anticipated by and extending into knowledge of the world. (p. 2579)

These definitions of language literacy and music literacy necessitate a model that incorporates the skills and knowledge for communication in addition to skills and knowledge about the content area. Figure 2.1 illustrates both essential components to literacy.

Left Side: Discipline Background

Discipline background describes the cultural, historical, and scientific context constituting the study of a literary or linguistic work (Bernstorff, 2016). These skills include understanding different patterns of reasoning and logic with the ultimate goal of processing information as a content expert. They also include understanding the culture and climate in which the authors worked as important components of literacy instruction (Fang & Coatoam, 2013; Urquhart & Frazee, 2012). In the language model, component categories include cultural context (see Kozulin, 2003), historical context (see Downey & Long, 2015), and scientific study
such as audiology, etymology, phonology, etc., see Hanauer, 2008). In the music model, parallel component categories include ethnomusicology (see Nettl, 1992), historical musicology (see Butt, 2002), and acoustics and theory, as well as music’s connection to other disciplines such as art, dance, and theater (see Parncutt, 2007).

Background knowledge is also represented in work by influential music pedagogues and philosophies. For example, Kodály, Suzuki, Jaques-Dalcroze, Orff, and Gordon all advocated that music learning involves more than simply reading notation (Chosky, Abramson, & Gillespie, 2001; Gordon, 2003). The Comprehensive Music through Performance (CMP) model also supports a holistic, multi-dimensional approach to music literacy that includes left-side and right-side components of the proposed literacy model (ILCMP, 2014). Moreover, the National Association for Music Education (NAfME, 2015) advocates for a comprehensive approach to music education that expands non-traditional paths to music literacy through the Core Arts Standards.

**Right Side: Communication Skills**

Communication skills reflect a more traditional view of literacy. Harris and Hodges (1995) defined literacy simply as “competence in reading and writing” (p. ix). McIntire (2007) also included listening and speaking in his definition. In the language literacy model developed for this study, speaking and writing are outputs that are demonstrated after cognitive processing. Listening and reading are inputs, processed during reception or by memory after reception (Ausubel, 2010). When transferred into the parallel realm of music, this definition suggests that music literacy is comprised of a student’s ability to read, compose, hear, and perform music (Mills & McPherson, 2015). Performing and composing are outputs, while, like the language
literacy prototype, listening and reading remain inputs. All communication skills are sub-components on the right side of the proposed language and music literacy models.

Note that reading and performing are not mutually dependent. That is, reading is not a prerequisite to performing and vice-versa (Mills & McPherson, 2015). Furthermore, the indicated communications skills are rarely learned in equal representation. Some non-Western cultures value rote learning (listening) over reading notation, and others value performing over creating (composing) new music. More discussion regarding non-Western cultures follows. Western cultural traditions, however, place enormous value on reading music notation, especially in the school environment. The emphasis on the reading, interpretation, and communication of printed music symbols is the *de facto* approach of most elementary and middle school band programs, as indicated by the content of method books and research regarding young ensemble rehearsal techniques (Cavitt, 2003; Colson, 2012; Heuser, 2011; Kratus, 2007; Sheldon, Balmages, Loest, & Sheldon, 2010). The advantages and shortcomings of this approach will be identified as the four communication sub-components are explored individually.

Speaking/performing and writing/composing are observable products of musical understanding. As outputs, they are based on an internal connection to prior knowledge. Speaking and performing rely heavily on the student’s prosodic development, which is the student’s capacity to “put words together into natural speech rhythms with intonation, inflection, and flow” (O’Herron & Siebenaler, 2007). Evidence shows that prosodic pitch changes, combined with reading speed, correlate with greater reading comprehension rates (Miller & Schwanenflugel, 2008).

Prosody depends on fluency. Armbruster and Osborn (2003) recommend that oral reading under the guidance of an instructor is the best means to practice expressive, fluid
prosodic reading or speaking. Similarly, rhythmic fluency is best developed through musical performance under the guidance of a music instructor (Hansen et al., 2014). Research shows that oral reading carries the same prosodic benefits as rhythm discrimination skills (Barlar, 2010; Colwell, 1994; Douglas & Williatts, 1994). These parallel prosodic reactions help indicate why music and speech share similar physiological reactions within the brain (Thaut, 2008).

Most schools emphasize writing skills early in the development of young children, beginning around five years old (Mills & McPherson, 2015). Writing skills help children increase their orthographic awareness while learning the conventions of grammar, spelling, and penmanship. The same benefits extend to music writing as a means for learning the conventions of music theory and notation. However, music-composing lacks the formalized pedagogy that predicates text writing (Parry-Jamieson, 2006; Tremblay-Beaton, 2015; Waller, 2010). Music educators, according to Waller (2010), should balance their music reading activities with composing activities to help reinforce coding fluency.

Composition differs from improvisation. Although both use a creative process, composition involves writing music ideas using some form of notation, iconography, or symbols, while improvisation involves the fluent performance of sounds within the parameters of the piece performed in real time (Aebersold, 2000; Larson, 2005). Many educators who write about the creative process treat them similarly, especially when exploring melodies by ear (Hickey & Webster, 2001; Priest, 2002). However, the compositional benefits of long-term recall and sharing of creative ideas among others are moderated by the extra time and effort required to scribe using the conventions of notation. Improvisation, on the other hand, generally requires the same amount of overall processing time, although time is more frequently invested in preparation for performance rather than the execution of performance (Larson, 2005). Outside of jazz
applications, these creative processes are seldom found in traditional band programs because they demand time and are difficult to assess (Heuser, 2011). As a result, composition and improvisation will likely continue to serve a subservient role to writing within the proposed literacy model.

Listening requires two steps to process information: decoding and interpretation/analysis. Listening skills are universally learned at an early age before the use of iconography or language symbols (Mills & McPherson, 2015). Young children begin aural decoding by developing their phonological awareness, which Hansen et al. (2014) define as the ability to identify and discriminate the characteristics of sound (frequency, intensity, time, and source). This ability can be represented through a variety of musical responses, including movement, singing, chanting, and clapping, although an observable response does not definitively indicate understanding (Ausubel, 2010). Studies have shown significant correlations among students’ ability to discern phonemes (language awareness) and pitch discrimination (music awareness) (Gromko, 2005; Lamb & Gregory, 1993; O’Herron & Siebenaler, 2007).

Once the stimulus is decoded, the listener must interpret and analyze it. This stimulus invokes a cognitive response, resulting in memory acquisition, problem solving, thoughtful contemplation, or the development of new ideas. The response is mediated by the student’s discipline background (Mills & McPherson, 2015), or the left side of the proposed literacy model. For example, a director may tell a trumpet player that she is sharp, and the musician, using the appropriate background knowledge, pulls out her tuning slide. Another example involves a musician who recognizes that the themes from the third movement of H. Owen Reed’s *La Fiesta Mexicana* are actually iterations of the same themes from the first two movements, and therefore makes stylistic adaptations so the themes are consistent.
Reading also requires decoding, followed by interpretation and analysis to process information. Reading begins after listening skills have developed in young children (Mills & McPherson, 2015). Whereas listening requires phonological awareness, reading requires orthographic awareness. Orthographic awareness consists of a child’s ability to identify and manipulate a grapheme, which is a printed symbol corresponding to an audible phoneme. Graphemes could consist of a single letter (b, g, t, s) or a group of letters (th, ch, sh). As students gain experience, they begin combining structures into words, phrases, and sentences (O’Herron & Siebenaler, 2007).

Orthographic awareness parallels a similar recognition process of music notation (Hansen et al., 2014). Graphemes in music equate to printed notes, rests, articulation marks, accidentals, dynamic marks, and other individual music symbols. As students gain music reading experience, they begin interpreting note patterns, motifs, phrases, and large form structures that help their performance become more fluid and more interpretive (McPherson & Gabrielsson, 2002).

Reading specialists have dedicated volumes to provide explicit strategies to aid student comprehension. Such strategies include think-alouds, visualization, coding, annotating, reciprocal teaching, and graphic organizers (Daniels & Zemelman, 2014; Hansen et al., 2014; Urquhart & Frazee, 2012; Vacca et al., 2011) and are explored in Appendix F. However, these strategies are typically only applied in a language literacy learning environment (Hansen et al., 2014). Since the evidence shows strong relationships between text literacy and music literacy, conventional methods of literacy instruction (i.e., direct instruction) may be severely limiting students’ potential as holistically literate musicians. If research dictates the viability of think-alouds, visualization, coding, annotating, reciprocal teaching, and graphic organizers for language comprehension, they certainly must also be viable for music notation.
Table 2.1 shows many of the common strategies explored by music educators Hansen et al. (2014) as well as three highly-cited content-area literacy texts currently available. Although the number of citations is not an indicator of its usage frequency in practice, the number of citations does indicate an interest in the content-area literature that may provide a starting point for developing an empirical study examining the relationship between text and music instructional strategies. Frequent strategies are explained in further detail in Appendix F.

Music educators must reevaluate their approach to music literacy. The incorporation of explicit textual reading strategies in music opens the door to new approaches to teaching music literacy. This is a first step toward the ultimate goal of valuing all aspects of literacy. In music, the conventional methods fall short of this goal because they rely heavily on notation, a staple of Western cultural traditions. A more holistic perspective also values non-Western cultural traditions, many of which diminish the importance of notation. An investigation of music literacy is incomplete without a comparison of Western and non-Western cultural traditions.

**Western versus Non-Western Traditions: Implications for the Model**

The music literacy model bears no assumptions for equal representation in any of its component features. Different cultures emphasize different aspects of literacy, by which a literate person in one culture may not be considered literate in another. While the proposed model serves as a prototype for literacy instruction, practical realities and cultural priorities prevent its full and equal implementation in most cultures. In many cultures, literate musicians are defined according to a different standard than Western cultural traditions might expect.

Western traditions originated from early European secular and sacred music practices in which status was awarded to the few who could interpret notation in its earliest
Table 2.1
Common Content Area Reading Strategies within Four Texts, 2011-2014

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic conversation</td>
<td>p. 71</td>
<td></td>
<td></td>
<td>p. 94, 165</td>
</tr>
<tr>
<td>Annotating</td>
<td>p. 121</td>
<td>p. 339</td>
<td></td>
<td>p. 170</td>
</tr>
<tr>
<td>Anticipation guides</td>
<td></td>
<td></td>
<td></td>
<td>p. 184</td>
</tr>
<tr>
<td>Coding</td>
<td>p. 125</td>
<td></td>
<td></td>
<td>p. 342</td>
</tr>
<tr>
<td>Concept circles</td>
<td></td>
<td></td>
<td></td>
<td>p. 169</td>
</tr>
<tr>
<td>Directed Reading-Thinking</td>
<td></td>
<td></td>
<td></td>
<td>p. 96</td>
</tr>
<tr>
<td>Exit slips/Admin slips</td>
<td></td>
<td></td>
<td></td>
<td>p. 222</td>
</tr>
<tr>
<td>Frontloading with images</td>
<td></td>
<td></td>
<td></td>
<td>p. 140</td>
</tr>
<tr>
<td>Graphic organizer</td>
<td>p. 147</td>
<td>p. 107</td>
<td></td>
<td>p. 244, 324.</td>
</tr>
<tr>
<td>KWL</td>
<td>p. 104</td>
<td></td>
<td></td>
<td>p. 206</td>
</tr>
<tr>
<td>Learning Log</td>
<td></td>
<td></td>
<td></td>
<td>p. 129</td>
</tr>
<tr>
<td>Metaphor and analogy</td>
<td></td>
<td></td>
<td></td>
<td>p. 341</td>
</tr>
<tr>
<td>Partner reading (performing)</td>
<td></td>
<td></td>
<td></td>
<td>p. 199</td>
</tr>
<tr>
<td>Password (vocabulary)</td>
<td>p. 115</td>
<td></td>
<td></td>
<td>p. 135</td>
</tr>
<tr>
<td>Post-it response notes</td>
<td>p. 118</td>
<td></td>
<td></td>
<td>p. 388</td>
</tr>
<tr>
<td>Prereading quiz</td>
<td>p. 107</td>
<td>p. 145</td>
<td></td>
<td>p. 94</td>
</tr>
<tr>
<td>Reading (performing) aloud</td>
<td>p. 98</td>
<td></td>
<td></td>
<td>p. 122</td>
</tr>
<tr>
<td>Reciprocal teaching</td>
<td></td>
<td></td>
<td></td>
<td>p. 161</td>
</tr>
<tr>
<td>Socratic seminar</td>
<td></td>
<td></td>
<td></td>
<td>p. 200</td>
</tr>
<tr>
<td>Summarization</td>
<td></td>
<td></td>
<td></td>
<td>p. 172</td>
</tr>
<tr>
<td>Think-alouds</td>
<td></td>
<td></td>
<td></td>
<td>p. 193</td>
</tr>
<tr>
<td>Turn and talk (reflection)</td>
<td></td>
<td></td>
<td></td>
<td>p. 335</td>
</tr>
<tr>
<td>Tweet the text (reflection)</td>
<td></td>
<td></td>
<td></td>
<td>p. 94</td>
</tr>
<tr>
<td>Vocabulary tree</td>
<td></td>
<td></td>
<td></td>
<td>p. 200</td>
</tr>
<tr>
<td>Vocabulary predictions</td>
<td></td>
<td></td>
<td></td>
<td>p. 197</td>
</tr>
<tr>
<td>Word wall</td>
<td></td>
<td></td>
<td></td>
<td>p. 143</td>
</tr>
</tbody>
</table>
forms (Treitler, 1982). Partially a product of the spread of Christianity in ninth and tenth century Europe, notation was valued by those seeking a reliable and standardized form of musical communication in the church (Burkholder, Grout, & Palisca, 2014). Notation became a status tool for the musically educated and the spiritually pure, while those who could not read notation were denied such status, regardless of their virtuosity.

Non-Western traditions, however, do not share the same linguistic foundations. Nettl (2015) observed, “In some cultures…music has been constructed in a way similar to language; in others it has been constructed rather differently, more influenced, one might guess, by dance, ritual, or emotion” (p. 310). These influences deemphasize the role of notation as a component of music-making. Ellingson (1992) and Kaufmann (1967) have explored many non-Western notational systems, noting that their function was mainly to preserve oral traditions, often serving as mnemonic aids to performers rather than as tools towards creating artistic structures. Therefore, the precision and virtuosity that emerged from Europe were not necessarily a priority in many non-Western cultures (Nettl, 2015).

Age: Implications for the Model

The instructional model that is traditionally used in American schools does not follow a balanced approach to the proposed music literacy prototype (see Figure 2.2), particularly among young performance ensembles (Bernstorf, 2016, Tremblay-Beaton, 2015). Ensembles are modeled after their professional counterparts, who use direct instructional techniques aimed at accurately decoding music notation. According to Tremblay-Beaton, “We teach the effective use of conventional tools instead of creative or alternative ways of achieving the tasks these tools were designed to execute” (p. 2579). Using a progressive approach, students may build a
musical understanding based on transforming ideas rather than replicating them. This includes attaining contextual knowledge as well as composing and listening skills. Conventional practices of music education tend to produce students that are missing a holistic and balanced approach to music literacy.

Understanding that decoding skills are components of the overall literacy model, research has explored how to teach young music students those skills more efficiently and effectively. The research shows, for instance, that children learn music best by moving from known melodies to unknown melodies to achieve technical mastery (instrumentally or vocally) in many disparate contexts (Mills & McPherson, 2015). Yet most instrumentalists in school settings are taught notation from the first or second lesson, failing to audiate or associate sounds with the technique required of the instrument (Schleuter, 1997). This process creates a pedagogical asymmetry between the internalization of pitch and rhythm and the external replication of pitch and rhythm. While students have an external referent for the technical manipulation of the instrument (which fingers to push down, for example), no such referent exists for rhythm. Therefore, many students who fail to learn sound before notation make rhythmic errors at a far greater rate than pitch errors (Dodson, 1983).

Research on the age and ability continuum of music literacy understanding has highlighted a disconnect between conventional practices and ideal practices (Tremblay-Beaton, 2015). Whereas the literature describes literacy as a comprehensive multi-dimensional phenomenon, practical descriptions of literacy strategies focus mainly on communication skills, especially reading. Tremblay-Beaton’s (2015) harsh criticism of traditional techniques is warranted based on the limited scope of current practices, as described by Cavitt (2003), Kratus (2007) and Heuser (2011). This disconnect highlights the need for more research exploring the
efficacy of non-conventional literacy practices, which includes a greater emphasis on listening skills and creativity (composition and improvisation) among beginning performance music students (i.e., band and orchestra students).

Investigating Relationships among Band/Music and ELA/General Knowledge

The following section examines several studies that have systematically explored the reciprocal relationship between music instruction and general knowledge. Preliminary research has suggested a positive correlation between band participation and literacy achievement. However, causation has yet to be determined. This literature review will explore the related research by identifying experimental, quasi-experimental, causal comparative, correlational, and phenomenological studies that are delineated similarly to the current study. Such studies focus on the relationship between band and ELA achievement in elementary or middle schools. The next section will focus on the relationship among all types of music instruction and all forms of generalized intelligence in elementary or middle schools. It includes related qualitative literature. The final section will examine the findings of meta-analyses on the relationship between music and literacy.

Experimental Evidence: Band and ELA Instruction

Causal evidence is often produced as a result of randomized experimental research (Cook & Sinha, 2006; Mertens, 2015). A search was conducted for randomized experimental research that investigated music reading in band and text reading in ELA with the following qualifications for inclusion:

- Literature required peer review
• Literature contained an analysis of music reading as a cause for a change in text reading ability or text reading as a cause for a change in music reading ability
• Literature contained systematic empirical inquiry
• Literature limited to the population of interest: elementary and middle school students
• Literature used band/non-band participation as an independent variable
• Literature employed a controlled pretest and posttest experimental design

Several electronic databases were searched from their inception to 2017: Educational Resource Information Clearinghouse (ERIC), Academic Search Complete, ArticleFirst, Google Scholar, and ProQuest Digital Dissertations. The search term music was combined with the search strings band AND (read or ELA or literacy or language) AND (effect or outcome or measure or evaluation). The search revealed no matching studies in the major educational databases and search tools presented thusfar.

**Non-Experimental Evidence: Band and ELA Instruction**

Quasi-experimental designs are similar to randomized experimental designs in structure and purpose except that participants are not randomly assigned to groups (Cook & Sinha, 2006; Mertens, 2015). These studies illustrate a pattern indicating a strong association between music and ELA. Neuharth (2000) performed a quasi-experimental study featuring a pretest-posttest non-equivalent control group design. Students self-selected into the experimental group by joining band in fourth-grade. He used fourth-grade scores as a covariate to determine that band students in eighth grade had made significantly greater gains in vocabulary, comprehension, linguistic mechanics, and science, $F(1, 103) = 4.073, p = .032$. However, they did not show significant gains in mathematics. He analyzed an eight-year time period involving five complete
Barlar’s (2010) quasi-experimental study investigated whether an intensive 14-week sight-reading instructional program significantly improved band students’ music sight-reading fluency and language reading fluency. Students (N = 53) with one to three semesters of playing experience were divided into two band classes according to the needs of their class schedule. Classes contained comparable ability levels. Barlar found no pretest differences between the experimental and control groups on reading fluency or on sight-reading ability, but she did find that high reading scores correlated with high sight-reading scores overall. Within both Neuharth’s (2000) and Barlar’s studies, students were not randomly assigned to each group, so unintended confounds could not be ruled out as a contributing to the results.

Causal comparative and correlational designs are similar in that they attempt to compare group characteristics that are latent or non-manipulatable (Mertens, 2015). Causal comparative designs focus on describing two or more group differences after an event has occurred, or ex post facto, while correlational designs focus more on describing the relationship between variables within a single group. Neither design indicates causation, as neither can “create groups that differ only within the limits of sampling error on both measured and unmeasured variables” (Cook & Sinha, 2006, p. 555).

Wachtel’s (2006) study, for example, could be described as a correlational design. He investigated whether a relationship existed among three literacy skill variables: verbal literacy (using standardized reading tests), note identification (using Music Ace software), and musical sight reading (using the Watkins-Farnum Performance Scale) among 30 fourth and fifth-grade students in band. He employed a Pearson product-moment coefficient and a Spearman rho
correlation, finding that there was a correlation between the Illinois Standard Achievement Test (ISAT) and Music Ace scores, but finding no correlations among the Iowa Test for Basic Skills, group assignment, or the Watkins-Farnum Performance Scale (WFPS). This project was completed in fulfillment of Wachtel’s master’s requirement and, as such, lacked theoretical constructs or a comprehensive methodological analysis. Despite being an otherwise informative and oft-cited study, these weaknesses reduce the external validity of Wachtel’s results. No causal comparative studies were found relating band instruction to ELA instruction.

Qualitative studies were searched using several electronic databases from their inception to 2017: Educational Resource Information Clearinghouse (ERIC), Academic Search Complete, ArticleFirst, Google Scholar, and ProQuest Digital Dissertations. The search term music was combined with the search strings band and (read or ELA or literacy or language) and (qualitative or interview or case). This search provided no qualitative studies exploring the relationship between band instruction and literacy instruction. However, two relevant studies did investigate the wider relationship among overall music and literacy using a qualitative design (Marrero, 2015; Shuck, 2005). They are discussed later in Chapter 2.

Experimental Evidence: Music and General Intelligence

Once the search for literature investigating the specific constructs of band and ELA achievement was completed, the search was widened to include all forms of music instruction and their effect on all forms of generalized intelligence. For example, a single study has investigated the impact of music instruction on literacy using an experimental design, performed by Costa-Giomi (1999). She found that randomly assigned low-income fourth-grade students improved their spatial abilities and their general cognitive skills after two years of private piano
lessons $F(3, 228) = 3.90, p = .01$, but the effects were temporary. She also found negligible effects on quantitative reasoning and verbal cognitive skills. She did conclude, however, that the improvements she observed were due to the experimental treatment and not due to chance. Costa-Giomi calculated omega squared effect sizes of .04 for spatial abilities and .02 for general cognitive growth, which Keppel (1982) indicated were low for the behavioral and social sciences.

Non-Experimental Evidence: Music and General Intelligence

Several non-experimental studies demonstrated an association between music and general intelligence. They are presented in chronological order starting in 1998, which is approximately when two meta-studies were published that summarized the previous research, (e.g., Butzlaff, 2000 and Winner and Cooper, 2000) . The meta-studies will be explored later in Chapter 2.

Wallick’s (1998) causal comparative study sought to examine the effects of a pull out string program on student achievement on several categories within the Ohio Proficiency Test (OPT). Two groups of students ($N = 296$) were equally divided into a string pullout-lesson group and a control group. Wallick used a paired samples $t$ test to investigate mean score differences between the groups and found a significant difference between string participants and non-participants on OPT reading scores. The difference was significant ($t = 2.127, p = .034$). A significant difference was also observed ($t = 2.003, p = .046$) between the citizenship scores of string participants ($M = 231.3, SE = 2.04$) with non-string participants ($M = 224.8, SE = 2.25$). Positive interactions were found regarding string participation and writing ($t = 1.270, p = .205$)
and string participation and mathematics ($t = .998, p = .319$), but the variability was too high to show statistical significance at the .05 level.

Kemmerer (2003) also used a causal comparative design to evaluate whether increased general music instruction in kindergarten through third grade improved standardized language, text reading, and music reading scores in fourth-grade as measured by the Terra Nova test and the Iowa Test of Music Literacy. She matched two school districts based on average economic status, total population, minority population, and mobility. From a small sample size ($N = 27$), Kemmerer found that additional instruction in music (17.5 minutes) did not correlate with higher language and reading scores on the Terra Nova test nor did additional instruction correlate with higher music reading scores on the Iowa Test of Music Literacy. She cited Green, Salkind, and Akey (2000) as determining that a power of .40 was strong enough to represent her findings. However, chances of a Type II error were high.

Gromko’s (2005) quasi-experimental study sought to determine whether music instruction was related to the development of kindergartners’ phoneme-segmentation fluency. She used four intact classrooms ($n = 43$) for her experimental group and four intact classrooms ($n = 60$) for her control group. The experimental group received 30 additional minutes of music instruction for four months. Using independent sample $t$ tests, she ascertained that the experimental group scored significantly higher on the test for phoneme-segmentation fluency ($t = -3.52, df = 101, p = .001$). This finding counters Kemmerer’s (2003) findings, illustrating the need for further research.

Johnson and Memmott’s (2006) causal comparative study examined over 4,700 students in a national sample of elementary and middle schools. Their research question asked if there was a relationship between the quality of the music program and standardized test scores and
whether geographical area impacted that relationship (e.g., South, East coast, Midwest, and Western portions of the United States.) The authors consulted experts to determine which band and orchestra programs belonged in the high-quality category and which belonged in the low-quality category. They then compared 1,119 students’ test scores from each category in the form of z-scores. Post-hoc tests from an analysis of variance (ANOVA) determined significantly higher test scores in higher-quality programs, but the effect sizes were small ($\eta^2$ ranged from .005 to .105).

Kinney’s (2008) correlational study also investigated the effect of music participation on academic achievement using several covariates to isolate influential factors. Notably, his study is one of few that incorporated parental involvement and mobility to determine influence. He retroactively investigated test scores of high school instrumental music students through grades four, six, and eight. A multivariate analysis of variance (MANOVA) revealed a significant correlation between socioeconomic status and academic achievement and band participation and academic achievement. He also found that higher achievement scores were present before students began music instruction. Kinney (2008) reported effect sizes in using $\eta^2$, which ranged from .04 to .07.

Huber (2009) used a causal comparative design to investigate whether a relationship existed between music participation and 14 variables associated with music instruction. The data were gathered through a survey, which was administered to a randomized selection of 300 music and non-music students in grades six, seven, and eight, and then checked for a correlation using a Pearson’s $r$ correlation coefficient. Huber found that music students did score significantly higher on ELA tests than non-music students.
Kurt’s (2010) correlational study used a two-way repeated measures ANOVA to investigate whether six music-related variables contributed to higher scores among eighth-grade band members ($N = 38$). Participants were measured on the Iowa Test of Basic Skills (ITBS) and the reading comprehension, vocabulary, and science subtests of the Measures of Academic Performance (MAP). He found that all band students improved their test scores regardless of SES, gender, grade point average, instrument, music involvement, and music motivation.

Baker’s (2011) causal comparative study investigated the differences between music participants and non-participants among 37,222 eighth-grade students on the Louisiana Educational Assessment Program (LEAP) English/Language Arts (ELA) and mathematics tests. Using a MANOVA procedure, he found that music students scored significantly higher than non-music students, and students who were excluded from the arts for additional test preparation did not score significantly higher. However, the results do not disaggregate the type of music class (band, chorus, or orchestra), nor do they consider the relative quality of the music instructor as a predictor of academic success.

Thornton’s (2013) study compared over 2,000 music participants in Pennsylvania with over 4,900 non-music students in grades 5, 8, and 11. The research question asked whether students who participated in band, chorus, or orchestra outscored their peers on the Pennsylvania System of School Assessment (PSSA). Using two-tailed $t$ tests, Thornton found that students involved in these music classes earned significantly higher standardized test scores than non-music students, thus suggesting that the resources and time spent on music instruction did not have a negative impact on English/Language Arts (ELA) and math test scores.

Two qualitative studies also investigated the relationship between music and general intelligence. Shuck (2005) conducted a qualitative case study design using surveys,
observations, lesson plans, interviews, and student work samples to ascertain the levels and frequencies of music integration into core subjects. Using this data, she investigated whether the influence of core subject areas had a reciprocal effect on music. The study focused on the music integration practices of 14 teachers at one public elementary school. She found that the interactions benefited students academically and teachers professionally. Awareness and training were paramount to the success of integration, and a much greater degree of professional development was needed to sustain and improve interactive relationships. This study supports related research and helps to refine the current study’s foundational theory that music and core subjects are intrinsically linked and are learned best in collaboration rather than isolation.

Similarly, Marrero (2015) interviewed six elementary music teachers, three principals, and three curriculum specialists to investigate how and why music teachers integrate literacy strategies in their classrooms, how non-music teachers value music and literacy integration, and what professional development is needed to enhance current practices. Marrero’s design replicates Shuck’s (2005) study, with the exception that Marrero only used interviews to collect data. She concluded that teachers tend to feel confident about their integrative abilities, yet still required explicit professional development.

Meta-Analyses

Three meta-analyses investigating the relationship between music literacy and text literacy were conducted in 2000 and 2008. Butzlaff’s (2000) study explored six experimental and 25 correlational studies, finding that the correlational studies demonstrate a positive association between music participation and literacy achievement ($t = 4.2, df = 23, p < .001$). The experimental studies, however, were inconclusive ($t = 1.06, df = 5, p = .34$). Effect sizes
were calculated and compared using Pearson’s $r$, which revealed that correlational studies produced a transformed mean $r$ of .17, 95% CI [.09, .24], while his experimental studies produced a mean $r$ of .18, 95% CI [-.21,.52]. Thus, the experimental study findings oppose the correlational study findings, which adds to the uncertainty of the relationship among music literacy and text literacy.

Winner and Cooper’s (2000) meta-analysis of research from 1950-1998 was not focused on music exclusively but instead on the arts as a whole. Examining 27 correlational studies, Winner and Cooper determined that associations between the arts and literacy were positively significant ($t = 6.36, df = 26, p < .001$). However, 24 experimental studies failed to indicate significant group differences ($t = 1.66, df = 23, p = .11$). Like Butzlaff’s (2000) study, Winner and Cooper found strong correlations between arts/music and academic achievement, but no conclusive indication of causation.

Finally, Standley (2008) explored 30 studies that specifically targeted the effect of music interventions on reading skills. She combined correlational and experimental studies into a single analysis and found that music interventions were positively associated with reading skills using Cohen’s $d$ ($d = .32, p < .001$), which is strong compared to meta-analyses of other (non-music) reading interventions. Disaggregating the results, she found that Pre-K ($d = .62$) and elementary students ($d = .25$) tended to benefit more than junior high students ($d = .00$). In conclusion, all three meta-analyses trend toward suggesting positive correlations between music instruction and literacy. However, they cannot confirm a causal explanation.
Conclusion

Even though critics of federal educational legislation, particularly NCLB and RTTT, lament increased testing and decreased instructional time, conditions were established that favored the development of the CCSS. These standards incentivized teachers to collaborate in innovative ways to find commonalities among ELA, math, and content areas like music. Although the assessment of those standards remains controversial (see Ravitch, 2014), the idea that complex understandings can be learned through literacy in all content areas remains laudable. This context gives new importance to the investigation of literacy and music instructional strategies.

Music education in the United States may be perpetually embroiled in a state of conflict: Should music be accessible to all or just to the talented? Should music be taught for its aesthetic beauty or for its pragmatic benefits? Should music instruction focus on performance skills alone or include untraditional strategies and assessments? Moreover, can text reading strategies mutually benefit both ELA and music? One study will not provide the answer; a body of experimental research (with randomized samples and controlled independent variables) is certainly warranted and would meaningfully inform the music education community about whether an amelioration between two disparate contexts is possible. It is the goal of the present study to provide evidence that will supplement existing literature with the proposition that content-area literacy techniques can transform music literacy instruction while re-establishing music’s unique and critical role within a balanced, holistic curriculum.
CHAPTER 3

METHODOLOGY

Introduction

The purpose of this study was to examine the reciprocal effects of music literacy instruction and textual literacy instruction among elementary band students in conventional and literacy-enriched contexts. Chapter 3 will present the methods that will be used to implement the study. The first section will present the research questions, hypotheses, and design. The next section will describe the participant sampling techniques and the experimental treatment procedures, which will be followed by the data collection and data analysis procedures. The final section will consist of the limitations of the study.

Research Questions

The following research questions concern textual literacy achievement:

1. How does conventional band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?

2. How does literacy-enriched band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?
3. Does literacy-enriched band instruction yield a different effect from conventional band instruction on textual literacy scores among elementary students?

The following research question concerns music literacy achievement:

4. Does literacy-enriched band instruction yield a different effect from conventional band instruction on musical literacy scores among elementary students?

**Hypotheses**

Hypothesis 1: Students participating in a conventional band environment will exhibit greater textual literacy growth than students who do not participate in band when controlling for pretest scores. In other words, H1: (μ1 > μ0) where 0 is the control group (non-band) and 1 is the conventional band group. The null hypothesis states that no difference exists between the groups or the control group outscores the conventional band group, or H0: (μ1 ≤ μ0).

Hypothesis 2: Students who participate in a literacy-enriched band environment will exhibit greater textual literacy growth than students who do not participate in band when controlling for pretest scores. This is represented with H2: (μ2 > μ0) where 0 is the control group (non-band) and 2 is the literacy-enriched band group. The null hypothesis states that no difference exists between the groups or the control group outscores the literacy-enriched band group, or H0: (μ2 ≤ μ0).

Hypothesis 3: Students in the conventional band group will exhibit different textual literacy scores than students in the literacy-enriched band group when controlling for pretest scores. The null hypothesis suggests that no difference exists between the group means.
Otherwise stated, H₃: (µ₁ ≠ µ₂) where 1 is the conventional band group and 2 is the literacy-enriched band group, while the null hypothesis is similarly coded H₀: (µ₁ = µ₂).

Hypothesis 4: Students in the conventional band group will exhibit different musical literacy scores than students in the literacy-enriched band group. The null hypothesis suggests that no difference exists between the group means. Otherwise stated, H₄: (µ₁ ≠ µ₂) where 1 is the conventional band group and 2 is the literacy-enriched band group, while the null hypothesis is coded H₀: (µ₁ = µ₂).

Research Design

This study followed an experimental design, which produced the ability to quantitatively measure the change of a dependent variable over time through observation (Mertens, 2015; Shadish, Cook, & Campbell, 2002). This study operated under a post-positivist paradigm, seeking to find confidence that a causal relationship among two variables existed while still recognizing the complexities inherent in observing social phenomena (Gall, Gall, & Borg, 2007; Shadish et al., 2002). According to Mertens (2015), No Child Left Behind legislation (NCLB, 2002) promoted the use of quantitative studies by creating “a political climate that supports the use of experimental or quasi-experimental designs, preferably with random assignment to groups” (p. 128). Therefore, this study aligns with an influential contingency of post-positivist policy experts and their methodological approaches toward systematic inquiry.

Quantitative studies investigating the relationship between music literacy and text literacy are not rare (Butzlaff, 2000; Kinney, 2008). The literature is replete with quasi-experimental, causal comparative, and correlational designs that explore the relationship between text literacy
and music literacy. These studies use non-randomized samples or uncontrolled experimental environments to observe and collect data (e.g., Baker, 2011; Elpus & Abril, 2011; Kurt, 2010).

However, randomized experimental studies are rare (Butzlaff, 2000; Kinney, 2008). To date, only one known study, performed by Costa-Giomi (1999), investigates the relationship between music literacy and text literacy, but it falls beyond the age range and discipline (band) of the present study. Unfortunately, the overall lack of randomized experimental research relating music literacy to text literacy means that much of the existing literature may be potentially biased or contaminated with spurious variables (Winner & Cooper, 2000).

Although qualitative designs investigating the relationship between music literacy and textual literacy are also rare, the qualitative literature does explore individual facets of literacy in isolation. For example, qualitative studies investigating innovative approaches to music literacy include Bazan’s (2011), Heuser’s (2011), and Misenhelter’s (2000), while qualitative studies investigating the integration of music and core-subjects include Marrero’s (2015) and Shuck’s (2005). Conversely, qualitative studies exploring innovative approaches to textual literacy are plentiful and include well-documented investigations on content area literacy (Cantrell, Burns, & Callaway, 2009) and disciplinary literacy (Fang & Coatoam, 2013; Shanahan & Shanahan, 2008).

The current study attempted to infer a causal relationship between music literacy and textual literacy while filling the quantitative-experimental gap in the literature. This was attempted by randomly assigning participants, thereby distributing confounds evenly across all groups. Follow-up studies will certainly be needed to explore various phenomena, contexts, and implications of the reciprocal effects of music literacy instruction and textual literacy instruction in greater detail. This study provided a structure and methodology for initiating such research.
Participants

The population of interest consisted of non-special education band students, male and female, within their first year of band instruction. Since there is no universal beginning grade level for formalized band instruction in the United States, the experimentally accessible population consisted of urban students in the fourth-grade, ages nine and ten. Care was taken to consider population validity (Mertens, 2015) by matching the accessible population to the population of interest as closely as possible.

The participating school was selected because it met several criteria: first, the school administered the Northwest Evaluation Association’s (NWEA) Measures of Academic Performance (MAP) test as part of its extant assessment framework. Second, the school began formal band instruction in grade five, which allowed an extra-curricular experimental group in grade four. No such ensemble existed in fourth-grade; therefore, the study did not impact an established extra-curricular or curricular program. Third, the school was willing to allow access and resources for the study, including permission, facilities, instructional materials, and instructional staff. Fourth, the band instructor at the school met researcher-established criteria to help increase external validity regarding instructional delivery, such as possessing at least five years of teaching experience (similar to Marrero, 2015), licensure by the state, and primary employment teaching students at the beginning level. Letters inviting band teachers to participate are included in Appendix A.

The selected school was an urban elementary school in Illinois that served grades K through six. It hosted a dual language program throughout all seven grade levels, so all documents and permission forms were provided in both English and Spanish. The average class
size was 23 students. Students benefitted from the consistency of having the same band instructor in grades four through eight since this director also taught band at the junior high level. Students used the same method book typically used by fifth-grade beginners at that school, *Measures of Success* (Sheldon, Balmages, Loest, & Sheldon, 2010).

All fourth-grade students who were administered the MAP were eligible for this study, acknowledging that adaptations may have been required for students with severe learning disabilities or impairments. Students who did not qualify for the MAP or who did not take the MAP, but who still wanted to learn a band instrument were included in the study if their parents so chose, with consideration given to the accommodations granted by a Section 504 plan or an Individualized Educational Plan (IEP). Since data were not collected from these students, their participation did not impact the study. Ultimately, no one in the pool of participants required accommodations so none were administered. One hundred seven students were invited to participate in the study. Based on Wachtel’s (2006) study under similar circumstances, a 65% participation rate was expected. A 40% participation rate was attained.

Parents and students were informed of the study using an introductory letter at the beginning of the investigatory school year (see Appendix B). The letter was accompanied by a parental consent form and student assent form (see Appendix C), and a demographic survey (see Appendix D). Additionally, the researcher and instructor presented a brief overview of the study to the students and staff to help generate interest and accelerate the return of the permission forms. The researcher provided incentives to students who returned those forms promptly, regardless of their consent or assent status. The incentives consisted of free ice cream from a local Wendy’s and McDonald’s franchise, which was donated to the school at no cost. All forms and incentives were approved by the Institutional Review Board at Northern Illinois University.
Several protocols were implemented to help mitigate common reasons for not signing up for band. The entire program was offered to students at no charge. Method books, music stands, pencils, and high-quality used instruments were purchased from a reputable local music vendor (see Appendix E for the budget). The demographic survey contained an inquiry about establishing transportation through carpool pools so that no one would decline the invitation based on ride availability. School officials worked to minimize extra-curricular events on lesson days so that students were not double-scheduled. All announcements and study documents were translated into Spanish due to the large population of Hispanic/dual language students. Lastly, assurances were made in the survey that individual responses would be kept confidential and unidentifiable.

Students who returned the permission forms were then assigned to one of three groups—a control group, a conventional band group, and a literacy-enriched band group—using a computerized random number generator. To maintain an ethical and equal delivery of band instruction, students in the control group had an opportunity to learn an instrument after the posttest. Therefore, no student was denied an opportunity to learn an instrument because they were placed in the control group. This embodied a modified switching replications design, since all students had access to instruction at different times, though the control group students were not reassessed after instruction. The experimental groups presented an evening concert performance as a culminating experience of their program. The control group presented a daytime concert performance for the school at the end of their instructional period, as evenings were not available so close to the end of the school year. The model for delivering the experimental treatment, collecting data, and providing equal access to instruction is presented in Figure 3.1.
Treatment

The privacy of the participants was “protected in that the data they provided was handled and reported in such a way that the data could not be associated with the research participants personally” (Mertens, 2015, p. 353). Students’ names were used to match data but were removed during the interpretation of the data and the reporting of the analysis, maintaining student confidentiality. The name of the participating school and community were also kept confidential. All electronic records were kept on a password-protected jump drive, and all paper records were kept in a locked file cabinet at the researcher’s home.

The salient qualities of each experimental condition were preserved as discreetly as possible to prevent experimental treatment diffusion. To accomplish this task, the instructor was given guidelines (see Appendix F) and training to help differentiate between the goals and strategies for each instructional paradigm. The instructor was then asked to keep each instructional paradigm discrete. The instructor was also required to keep a log of activities and lesson plans to help maintain treatment fidelity (see Mertens, 2015).

The conventional band paradigm consisted of performance-based instruction in small group lessons and in full group rehearsals. Conventional band instruction is the overwhelming model used by band directors in the United States (Heuser, 2011). At the beginning stages of band instruction, this paradigm featured continuous repetition and muscle memory to help students make a kinesthetic connection between the way a note felt and the way it appeared on the page. Conventional techniques included clapping, singing, rote teaching, trial and error practicing, chanting, repetition, chunking (dividing material into its smaller component parts), teacher/peer modeling, verbalizing note names, verbalizing fingerings, and audiation. Students
Start

Study information/assent/consent forms sent home to all participants’ parents/guardians

Do students have permission to participate?

yes

Text literacy pretest (November MAP Test)

Attrition: Refusal Affordability* Transportation* Scheduling
*attempted to mitigate

Control

Students randomly divided into 3 groups

Literacy-Enriched

Conventional

Phase 1: Treatment (14 classes)

Phase 1: Treatment (14 classes)

Posttest (Apr. MAP)

Posttest (Apr. MAP)

Posttest (Watkins-Farnum)

Posttest (Watkins-Farnum)

A

B

C

Study complete. Data not included.

no

(continued on following page)
Figure 3.1. Procedure for experimental treatment, data collection, and equal access to instruction.
were also allowed to complete any of the non-performance activities explicitly suggested in their method book. Six lines asked students to write in note names, two lines asked students to draw a clef, and one asked students to conduct. Any additional markings or activities students initiated on their own accord were permissible if the student was not specifically prompted to do so by the instructor. These strategies were used both in small group lessons and in large group rehearsals involving heterogeneous instrument groups.

The literacy-enriched band paradigm consisted of all of the strategies outlined in the conventional paradigm with the addition of several specific reading strategies borrowed from content literacy and disciplinary literacy instructional guides, such as Daniels and Zemelman (2014), Urquhart and Frazee (2012), Vacca et al. (2011), and Hansen et al. (2014). These strategies included coding, annotating, think-alouds, and frontloading with images, and are described more thoroughly in Appendix F. Strategies were applied both to music reading and text reading (e.g., new musical concepts introduced in textual forms in the method book). Time did not allow for the inclusion of all applicable strategies in Appendix F, but it did allow for the inclusion of 10 discreet strategies. The application and timing of these strategies were coordinated with the fourth-grade reading teachers to help maintain a parallel consistency between the literacy-enriched band students and fourth-grade classes as a whole. A transcription of the coordination meeting is presented in Appendix G.

One literacy-enriched strategy was already regularly performed in the conventional paradigm—the read aloud. The musical version of the read aloud features the teacher performing a musical piece with stylistic fluency as a model to which students may aspire. Unlike rote teaching, the read aloud typically does not come from student material (Daniels & Zemelman, 2014). Regardless, this is the only explicit literacy-enriched strategy that is routinely
employed in the conventional setting (Henninger, Flowers, & Councill, 2006), and therefore was not denied to students in the conventional group.

The frequency of literacy-enriched strategies was calculated after the treatment was delivered (see Table 3.1). The most-used strategy was turn and talk, where students were directed to diagnose or interpret a musical issue with a peer. Some strategies were used two or more times, including vocabulary predictions, think alouds, summarization techniques, and frontloading with images. Each strategy is described and developed more fully in Appendix F.

Table 3.1
Frequency of Explicit Strategies Used in the Literacy-Enriched Group

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn and talk</td>
<td>5</td>
</tr>
<tr>
<td>Frontloading with images</td>
<td>3</td>
</tr>
<tr>
<td>Summarization</td>
<td>3</td>
</tr>
<tr>
<td>Think aloud</td>
<td>3</td>
</tr>
<tr>
<td>Vocabulary predictions</td>
<td>2</td>
</tr>
<tr>
<td>Annotations</td>
<td>1</td>
</tr>
<tr>
<td>Codes</td>
<td>1</td>
</tr>
<tr>
<td>Read aloud</td>
<td>1</td>
</tr>
<tr>
<td>Post-it response notes</td>
<td>1</td>
</tr>
<tr>
<td>R-D-W</td>
<td>1</td>
</tr>
</tbody>
</table>

To strengthen the internal validity of the study, a single band instructor taught both the conventional and literacy-enriched band classes. Additionally, the random selection of participants ensured roughly equal representation of students across all fourth-grade classes.
This mitigated variances due to the type of instruction beyond the experimental treatment.

Furthermore, treatment fidelity was ensured through the use of training, common instructional materials, classroom observations by the researcher, specific strategy protocols, and an instructor log where daily lesson plans were kept (see Appendix F).

For the sake of replicability, Table 3.2 provides a schedule representing the experimental treatment. Lessons and full band rehearsals were held in the cafeteria. As one group was engaged in their band lesson, others silently read a book or completed their homework.

Homework sessions received equal time among all groups. Percussion lessons were held before school once a week, while woodwind lessons, brass lessons, and full band rehearsals were held after school once a week. The conventional group met on Mondays and the literacy-

Table 3.2

Scheduling Template for Experimental Groups

<table>
<thead>
<tr>
<th>Time</th>
<th>Conventional Band: Mondays 14 Sessions</th>
<th>Literacy-Enriched Band: Wednesdays 14 sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Woodwinds</td>
<td>Brass</td>
</tr>
<tr>
<td>7:20-7:40</td>
<td>7:20-7:40</td>
<td>Small Group Lesson</td>
</tr>
<tr>
<td>3:00-3:30</td>
<td>3:00-3:30</td>
<td>Small Group Lesson</td>
</tr>
<tr>
<td>3:30-4:15</td>
<td>3:30-4:15</td>
<td>Full Ensemble</td>
</tr>
<tr>
<td>4:15-4:45</td>
<td>4:15-4:45</td>
<td>Small Group Lesson</td>
</tr>
</tbody>
</table>
enriched group met on Wednesdays. The treatment duration consisted of 14 thirty-minute lessons and 14 forty-five-minute full band rehearsals. Students participated in one lesson and one rehearsal each week over the course of 17 weeks to accommodate winter break and major holidays. This schedule was similar to Barlar’s (2010) 13 week, 45 minutes per week music intervention and Gromko’s (2005) 14 week, 30 minutes per week intervention.

Data Collection

This study utilized two data collection instruments: the Northwest Evaluation Association’s (NWEA) Measures of Academic Performance (MAP) to evaluate textual reading growth, and the Watkins-Farnum Performance Scale (WFPS) to evaluate musical achievement. Whereas all fourth-grade students took the MAP test, only students in the experimental (band) groups took the WFPS. A description of the MAP test will precede a description of the WFPS test. The data collection strategies and their alignment to the research questions are illustrated in Figure 3.2.

Description of the MAP

The MAP testing process consists of benchmark, or interim, assessments that are used by teachers to provide differentiated instruction, measure student growth, and compare progress to other students, classes, and schools (Cordray, Pion, Brandt, Molefe, & Toby, 2012). Data from the MAP were collected using a single-factor multiple-treatment design in which the researcher created a control group and two experimental groups (Mertens, 2015). Care was taken to
Figure 3.2. Alignment of data collection, data analysis, and research questions.
maintain equal sample sizes among groups by anticipating a 10% experimental group dropout rate and frontloading the experimental groups with extra participants. The pretest was administered to all fourth-grade students in November and the posttest was administered to all fourth-grade students in April.

The MAP was a computer-based, criterion-referenced adaptive test that helps schools diagnose student deficiencies and measure student growth in reading, language usage, mathematics, and science (NWEA, 2016). Test items became more or less difficult in response to students’ performance ability, adapting differently for correct or incorrect answers. Once the test was completed, scores were immediately available.

The scope of the present study delimited the investigation to the MAP reading test and its component subtests: literature, vocabulary acquisition and use, and informational text. (NWEA, 2017a). Despite the disaggregation of the reading test into those three subcategories, the test did not measure overall literacy, nor did it measure reading as a whole. The conceptual model presented in Chapter 2 suggests that the overall construct of literacy is much too broad to be distilled into the reading tests that the MAP provides. However, the MAP does show valid and reliable evidence of specialized reading growth, which was useful to the research questions posed.

The MAP reported achievement in the form of Ready for Instruction Today (RIT) scores, which were synonymous with Rausch Units (NWEA, 2017b). RIT scores were continuous variables that indicated a student’s zone of proximal development, regardless of age or grade level. The overall reading RIT score was an average of the three reading subsections (or subtests). According to the NWEA (2015), fourth graders typically earned a RIT score of 198.2
(SD = 15.53) at the beginning of the year and a score of 205.9 (SD = 14.92) by the end of the year.

The literature subtest specifically measured how well students could find key ideas and comprehend literary texts by predicting, inferring, and summarizing information (NWEA, 2017a). Additionally, this subtest investigated students’ ability to identify and evaluate the structure of a literary text. Structure included plot, setting, theme, and literary devices (Andren, 2010). The vocabulary acquisition and use subtest specifically measured how well students could identify contextual clues, understand word relationships, and decode unknown words. It also tested student knowledge of common vocabulary words. Lastly, the informational text subtest measured the same attributes as the literature subtest, but for non-literary works. Accordingly, this subtest also investigated how well students could identify bias and draw appropriate conclusions.

Validity and Reliability of the MAP

Reliability is defined by the American Educational Research Association (AERA, 1999) as “the degree that true scores are free from errors of measurement” (p. 93). MAP items are calibrated by the NWEA using a multiple step process: item development, pilot testing, field testing, and fine calibration to determine item difficulties (NWEA, 2009). Items are continuously monitored for bias, relevance, and clarity. Unusual items are identified using a point-biserial correlation index and an adjusted Root-Mean-Square Fit (RMSF) index, which compares student performance on a single test item to their performance on the overall test. The NWEA (2009) uses an algorithm to determine marginal reliability, which generates coefficients between .89 and .96 and are considered highly consistent (AERA, 1999). Test-retest
correlations, measured via several longitudinal norms studies, indicate Pearson coefficients between .79 and .94 for all students except second graders (NWEA, 2009). These results are also considered highly consistent (AERA, 1999).

Validity is defined by the AERA (1999) as “the degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of the test” (p. 184). The NWEA (2009) provided several indicators that the MAP assessments were appropriate tools for measuring student growth in literacy, including the multi-step item development process ensuring content validity, positive test modality comparisons, and training in test score interpretation. Several concurrent validity studies have also shown direct correlations between state standardized assessments and MAP scores, which ranged from .84 to .86 for fifth-grade students taking the MAP reading test. These correlations illustrated that MAP results were consistent with other tests used for similar purposes (NWEA, 2009).

Description of the Watkins-Farnum Performance Scale

The Watkins-Farnum Performance Scale (WFPS) (Watkins & Farnum, 1954) was applied via a one-shot case study design to investigate the differences between the effects of conventional band instruction versus the effects of literacy-enriched band instruction on musical ability. During this test, students performed their instrument to demonstrate a subset of their music literacy capabilities. Referencing the model described in Chapter 2, the WFPS assessed two components of comprehensive music literacy: reading and performance. The WFPS did not address writing or listening, nor did it address aspects on the left side of the model such as musicology or acoustical sciences. Despite these limits, data illustrating reading and performing
Capabilities formed a foundation to building an empirical base of research regarding the effects of conventional and literacy-enriched instructional techniques.

The WFPS (1954) protocols specifically define how a student’s playing performance is to be assessed objectively within eight categories: pitch, time/rhythm, time change, expression, slur, rests, holds/pauses, and repeats. The instructor grants point values based on the number of correct measures played. Internal validity is robust, as the test produced rank-order correlations from .66 to .91, depending on the musical instrument (Watkins & Farnum, 1954; Stivers, 1972), and has been used effectively in several applications assessing technical musical sight-reading ability (such as Barlar, 2010; Belfast, 2013; Elliot, 1982; Gromko, 2004; Wachtel, 2006).

The test begins when the student attempts to perform the first of 14 exercises, arranged progressively from easy to difficult (Watkins & Farnum, 1954). An assessor codes errors using eight error protocols: pitch, time, change of time, expression, slur, rests, holds and pauses, and repeats. Only one error may be counted per measure. The student continues playing until he/she performs zero correct measures in two consecutive selections. The final score is the overall number of correct measures played. The test represents a small component of the overall literacy model, so the results were interpreted with that limitation in mind.

Validity and Reliability of the WFPS

The test was developed by Watkins and Farnum (1954) to address invalid and unreliable music assessment practices prior to 1954. By creating clear constructs and specific protocols, the test became a widely-used, objective instrument to assess examinees’ music reading and playing skills. Watkins performed validity tests on the Watkins-Farnum scale by using rank-order correlations. In these tests, Watkins ranked students by ability, administered the test, and then
ordered them by test score. The correlation coefficients vary by instrument, but the mean coefficient was .81 and the standard deviation was 0.05, indicating that the test was a valid assessment of relative playing ability.

Reliability within the current study was measured by investigating the internal consistency of the observed scores. The tests were audio recorded by the instructor with an iPad and re-evaluated by a second assessor. Cronbach’s alpha was used to measure the internal consistency between raters. Inter-rater reliability was found to be very high ($\alpha = .953$). The targeted internal consistency range was from .75 to 1.00 (Dimitrov, 2013).

Data Analysis

Descriptive and inferential statistics were used to analyze the data collected in this study. Descriptive statistics were used to describe group characteristics on a single variable such as the mean and standard deviation. Inferential statistics were used to determine the probability that an effect, not chance, caused a difference among groups (Dimitrov, 2013; Mertens, 2015).

The descriptive statistics were largely based on the demographic survey, which determined how closely the sample reflected the population by investigating the gender, ethnicity, and socio-economic status of the participants (see Appendix D). Costa-Giomi (2012) identified several studies that demonstrated the effect of SES and family characteristics (e.g., mobility and single parents) on academic achievement among music students, including several studies mentioned in Chapter 2 (such as Baker, 2011; Huber, 2009; Kinney, 2008; Kurt, 2010). Costa-Giomi’s findings necessitated an investigation of these characteristics within the current study. Therefore, the sample’s mean demographic values were compared to state and national averages to determine the representativeness of the sample and equality within each experimental
group, but those data were not used to change random group assignments (i.e., proportional stratified sampling). Additionally, descriptive data were used to help confirm the assumptions for the statistical tests. The mean, standard deviation, skew, and kurtosis were investigated statistically and visually (via regression lines and scatter plots) to determine whether the error terms associated with the dependent variable met the assumptions of normality, whether the regression coefficients were independent, and whether the data were homoscedastic.

Inferential statistics were used to analyze the data from the MAP reading test (NWEA, 2016) and the Watkins-Farnum (1954) test. The MAP produced continuous data from three groups: a control group, a conventionally-taught band group, and a literacy-enriched band group. Data were analyzed as RIT units without transformation. The Watkins-Farnum Performance Scale produced continuous data from only two groups: the conventionally-taught band group and the literacy-enriched band group. The analysis procedures from each data collection instrument will be considered next.

Analyzing the Measures of Academic Performance (MAP)

The MAP test produced categorical data that were analyzed against one factor at three levels—a control group, a conventionally taught band group, and a literacy-enriched band group. Pretest scores were also available, as were the participants’ demographic data, which provided several statistical options for data analysis. Those options will be presented along with a justification for ultimately choosing an analysis of covariance (ANCOVA).

Statistical models are associated with research question types, such as associations and correlation (e.g., Pearson’s Correlation), prediction and relationships (e.g., linear regression), group differences (t tests or ANOVA), or reliability testing (Cronbach’s alpha or Cohen’s kappa,
Laerd Statistics, 2017). Since the primary purpose of this study was to compare means between three groups, the analysis of variance (ANOVA) family of tests was the best fit.

However, an ANOVA does not adjust the dependent variable for the covariance of the pretest scores. If pretest and posttest scores are available, one option is to analyze the gain score, which is the difference between the posttest and the pretest scores. Experts disagree on whether a one-way repeated measures ANOVA using the gain score as a dependent variable (called Gain Score Analysis, or GSA) is a better fit for a study than using a one-way ANCOVA with the posttest scores as a dependent variable and pretest scores as a covariate. The resulting discrepancy is known as Lord’s paradox (Lord, 1967). Knapp and Schafer (2009), von Breukelen (2006), and Wright (2006) claim that the ANCOVA is generally advantageous because gain scores may not be reliable, the ANCOVA produces greater power, and the gain scores correlate negatively to the pretest, particularly when pretest data are skewed. However, this is refuted by Allison (1990) and Edwards (2001), who suggest distortions are caused by other factors. Nonetheless, the two approaches answer different research questions. The GSA seeks to find the effect of the treatment on change from pretest to posttest, while the ANCOVA seeks to find the effect of the treatment on the posttest that cannot be predicted from the pretest (Dallal, 2005; Grace-Martin, 2013; Knapp & Schaffer, 2009). Considering this evidence, the ANCOVA was chosen for this study with the pretest scores serving as the covariate to adjust the posttest scores according to pretest differences.

Additional ANOVA options were presented for consideration. A Factorial Repeated Measures ANOVA and a Multivariate ANOVA both had the ability to investigate the primary interaction of interest (i.e., the group vs. means difference) in addition to investigating interactions of demographic variables to indicate the presence of influential confounds, such as
Factorial methods, however, required a larger sample size than was available for this study to establish sufficient power. Moreover, “MANOVA is substantially more complicated than ANOVA. There are several important assumptions to consider, and there is often some ambiguity in the interpretation of the effects of IVs on any single DV. Further, the situations in which MANOVA is more powerful than ANOVA are quite limited” (Tabachnick & Fidell, 2007, p. 244). These options thus lacked parsimony and reliability for the data collected.

Similarly, a hierarchical multiple regression (HMR) was considered. Using this model, the input of independent variables in hierarchical blocks allowed an opportunity to investigate and predict the effect of each block individually. Like the factorial design, the sample lacked the heterogeneity and sample size to represent each potential interaction with validity. Additionally, the blocks in this study were difficult to define and measure, since other than gender and teacher, the data for SES, parent status, and ethnicity were skewed, greatly reducing generalizability (Shalev, 2007). Therefore, HMR was not utilized in this analysis.

Hierarchical linear modeling, or HLM, was also considered as an alternative to General Linear Modeling (GLM), which includes the one-way ANCOVA. HLM applies to randomly selected groups within a larger set of groups (Tabachnick & Fidell, 2007), such as five schools randomly selected from a set of 100 possible schools (Huta, 2014). HLM requires a substantial sample size at each level of the model. In contrast, GLM applies to groups that had no alternative from which to choose, and thus no group randomization, which matches the research design presented herein. Therefore, GLM was the appropriate choice.

As a result, in the interest of parsimony, the demographic data were analyzed independently from the ANCOVA using t tests for the dichotomous variables (gender, SES, and
parent status) and a one-way ANOVA for the classroom teacher assignment. If any demographic variables were found to correlate with the posttest scores, they would have been incorporated as covariates or entered into a multiple regression analysis. These data were not automatically entered into the ANCOVA as additional covariates because the loss in the degree of freedom for each covariate would reduce power more than decreasing the sum of squares for error would raise power (Tabachnick & Fidell, 2007).

**Administration of the ANCOVA**

The one-way ANCOVA was administered to data from the overall reading test as well as its component subtests (literature, vocabulary acquisition and use, and informational text). It produced two sets of results for each test (see Figure 3.2). The first was the omnibus result, which measured the total interaction of all three groups but did not indicate the relative strength of each group interaction. The second set of results consisted of the planned contrasts which measured each group interaction individually. The omnibus procedures were necessary to produce covariate-adjusted mean post-test values, but the contrasts provided specific information about each group effect after the adjustment of the dependent variable by the covariate (Field, 2013). The contrasts were selected *a priori* to avoid losing statistical power with post hoc tests.

Research question 1 investigated the difference between the control group (group 0) and the conventionally taught band group (group 1). The data of interest were the planned contrast results derived from the overall reading test, the literature test, the vocabulary acquisition and use test, and the informational text test. A Bonferroni correction was applied to the contrasts to avoid an error of multiplicity (increased Type I error rate due to multiple investigations).
Minimal power was established at .80. With consideration to the hypothesis and statistical method used, the study required a sample size of 159 to show an effect size of $d = .50$, which was deemed practically significant by Rossi and Wright (1977). Due to the limitations (costs, availability, and access) of obtaining such a large sample size, a larger effect of $d = .75$ was calculated to require a sample size of 72, which was realistic according to prior literature (see Wachtel, 2006). It was therefore determined that if the sample size were to drop below the 72-participant threshold, the alpha (Type I error) rate would be raised to from .05 to .10 to help maintain minimal power. This decision is acceptable (Nuzzo, 2014) due to the result of a lack of related experimental literature, the arbitrary nature of alpha = .05, the naturally occurring variance of social behaviors, and the association of low $p$ values with sample size, not effect. Cook and Sinha (2006) also indicated the acceptability of raising the alpha rate under certain conditions:

Use of the statistical criterion of $p < .05$ is now standard and this may lead researchers to define program efficacy so stringently that they overlook some positive program effects. However, this trade-off reflects convention and is not intrinsically tied to experimentation. Researchers can use different alpha rates and describe findings that do not meet less stringent statistical criteria, so long as their decisions are explicitly justified and they present sufficient data to allow readers to compute different standards. (p. 559)

Thus, the alpha level and corresponding critical $p$ values were not chosen arbitrarily.

The null hypothesis stated that the control group (0) would be equal to or greater than the conventional group (1) after adjusting for the covariate (pretest score), or $H_0: \mu_1 \leq \mu_0$. The alternative hypothesis suggested that the conventional band group would have a higher mean posttest score after adjusting for the covariate (pretest score) than the control group, i.e., $H_1: (\mu_1 > \mu_0)$. Since the alternative hypothesis predicted a positive difference indicating growth, a right-sided one-tailed alpha level was used.
Research question 2 concerned the interaction of the control group (group 0) with the literacy-enriched band group (group 2). Similar to research question 1, the results of interest were the planned contrasts from all four ANCOVA procedures (overall reading, literature, vocabulary acquisition and use, and informational text). A Bonferroni correction was applied to all four \( p \) values to prevent an inflated alpha due to the multiplicity of procedures. Minimal power was established at .80 using the same effect sizes and sample sizes as research question 1, including an adjusted alpha level of .10 from .05 if the sample size dropped below 72.

The null hypothesis stated that the control group (0) would be equal to or greater than the literacy-enriched group (2) after adjusting for the covariate (pretest score), or \( H_0: \mu_2 \leq \mu_0 \). The alternative hypothesis suggested that the literacy-enriched band group would have a higher mean posttest score after adjusting for the covariate (pretest score) than the control group, i.e., \( H_2: (\mu_2 > \mu_0) \). Since the alternative hypothesis predicted a difference indicating growth, a right-side one-tailed alpha level was used.

Research question 3 concerned the interaction of the conventional group (group 1) with the literacy-enriched band group (group 2). Planned contrasts from all four tests were again used to investigate the research question, and a Bonferroni correction was applied to the \( p \) values due to the multiplicity of procedures. Minimal power was established at .80 using the same effect sizes and sample sizes as the previous questions.

The null hypothesis stated that there was no difference between group means (\( H_0: \mu_1 = \mu_2 \)), while the alternative hypothesis stated that there was a difference between conventional and literacy-enriched band groups (\( H_3: \mu_1 \neq \mu_2 \)). Hypothesis 3 indicated no directionality, as previous research was inconclusive determining whether conventional band and literacy-enriched band instruction produced different mean scores on the MAP. Therefore, this
procedure required a two-tailed alpha level and a higher alpha rate of .10 if the sample size dropped below 72.

An ANCOVA requires ten assumptions to optimally fit the data (Field, 2013; Laerd Statistics, 2015). The first four assumptions were related to the design of the study. The assumptions require that the study contained 1) a single, continuous dependent variable (posttest scores), 2) a single, categorical factor (group assignment), 3) a single, continuous covariate variable (pretest scores), and 4) participants that did not belong to more than one group (independence of observations). The design of the study ensured that these assumptions were observed for all four ANCOVA procedures (overall reading scores, literature subtest scores, vocabulary acquisition and use subtest scores, and informational text subtest scores).

The remaining six assumptions were tested statistically after each ANCOVA was administered. The results are included in Chapter 4. They include the assumption of linearity between the covariate and dependent variable, homogeneity of regression slopes, normally distributed residuals on the dependent variable across the factor, homogeneity of variances, and no residual outliers (Dimitrov, 2013; Field, 2013; Mertens, 2015).

Analyzing the Watkins-Farnum Performance Scale (WFPS)

The Watkins-Farnum Performance Scale (Watkins & Farnum, 1954) measured the musical reading and playing abilities of the examinees chosen for the two band-experimental groups. Since the control group did not participate in band instruction, they were not administered this test. The WFPS consisted of protocols for assessment and 14 progressively ranked musical selections. The protocols included instructions for counting errors in pitch, time/rhythm, time change, expression, slur, rests, holds/pauses, and repeats.
For example, the protocols specified, “If the student strikes the wrong pitch when attacking a note but correctly fingers it and immediately adjusts the lip to the correct pitch without re-tonguing the note, no error is to be counted” (p. 6, emphasis in original). Similarly, “A sustained note must be held within one count of the correct beat. Thus, a whole note held for three full counts is marked wrong. If held for three counts and a little more it is considered right” (p. 7, emphasis in original). The musical selections were ordered from a beginning level to advanced level. Students continued attempting the test until they scored zero points on two consecutive selections.

Administration of the Independent Samples t Test

Research question 4 examined the interaction of the conventional group (group 1) with the literacy-enriched band group (group 2) on the Watkins-Farnum Performance Scale. The WFPS generated numerical scores based on the number of measures performed correctly by the examinee. The independent variable was the method of instruction (i.e., conventional band instruction or literacy-enriched band instruction). The dependent variable was playing and reading ability as measured by the WFPS. Since the scores consisted of interval level data and since the independent variable had two categories, data were analyzed using an independent samples t test.

The null hypothesis stated that there was no difference between group means (H0: μ₁ = μ₂), while the alternative hypothesis stated that there was a difference between conventional and literacy-enriched band groups (H₄: μ₁ ≠ μ₂). Like hypothesis 3, hypothesis 4 indicated no directionality, as previous research was inconclusive determining whether conventional band and
literacy-enriched band instruction produced different mean scores on the WFPS. Therefore, this procedure also required a two-tailed alpha level.

A targeted power of .80 was established, which used the statistical model and hypothesis to provide a suggested sample size of 144 when the effect size $d = .50$, which Rossi and Wright (1977) considered clinically significant. Considering that obtaining such a sample size was prohibitive, an alternative sample size was computed to be 58 if the effect size was increased to $d > .75$. Similar to the ANCOVA analysis, the alpha rate was to be raised if the sample size fell below the 58-participant threshold.

Six assumptions of the $t$ test were evaluated. Three related to the study design, which required that: 1) the dependent variable contained continuous data, 2) the independent variable contained categorical data in two groups, and 3) the observations were independent of each other. These assumptions were met. The final three assumptions were assessed after the $t$ test was administered, and are discussed in Chapter 4. They include the assumption of no outliers, residuals normally distributed across all levels of the independent variable, and homogeneity of variances.

Building Cause and Effect Arguments

Research questions 1 and 2 contained generalized causal aspirations. Determining cause and effect relationships is both philosophically and statistically problematic. The primary argument against randomized experiments as a means toward generalized causal inferences is that experiments are globally aspired yet locally defined (Shadish et al., 2002). In other words, external validity is often sacrificed in favor of internal validity. Furthermore, causal relationships are rarely direct, linear relationships. Instead, they “entail a systems model of
interrelationships that is more akin to intersecting pretzels than to the experimenter’s simple arrow from A to B” (Cook & Sinha, 2006, p. 557). Perfectly heterogeneous sampling helps overcome confounding variables, but formalized and balanced probability sampling methods (i.e., simple random sampling) are not viable nor practical in the social sciences. In other words, no sample is entirely representative of the general population. However, Shadish et al. (2002) noted, “Practicing scientists routinely make causal generalizations in their research, and they almost never use formal probability sampling when they do” (p. 24). As a result, Shadish et al. (2002) have studied scientific causal inferences and proposed a process for evaluating generalized causal claims based not just on statistical tests, but on five principles for crafting generalized causal inferences.

Shadish et al. (2002) begin by asserting that causes are generally created by an “inus” condition (see Mackie, 1974), meaning the condition was an “insufficient but non-redundant part of an unnecessary but sufficient condition” (p. 62, italics in original). In the case of an educational treatment, a cause may be insufficient because it could not cause the effect without other (often latent) conditions present, such as supportive households, safe facilities, or electronic resources. The cause may be non-redundant because it was a unique addition to the curriculum. It may be unnecessary because other combinations of events may have also led to the same result. The combination, however, was sufficient for the event or result to occur. When arguing for causation, treatments are generally accepted as “inus” conditions because knowing the types and relationships of all moderating and mediating variables is not possible in behavioral research (Shadish et al., 2002). For this reason, causation is always a probabilistic approximation, not an absolute determination.
Effects are best understood with a counterfactual approximation. True counterfactuals require that the same participants both receive the control and treatment conditions, which is not possible. Thus, counterfactuals are approximated by randomized sampling. The reliability of the experiment rests heavily on the heterogeneous variability of the chosen sample, such that pre-existing, probabilistic differences of all confounds in the sample are spread equally across all groups. Most social science experiments use purposive sampling strategies that attempt to either sample typical instances or heterogeneous instances depending on the study design (Cook & Campbell, 1979). Formal sampling of this type helps ameliorate the gap between a local design and a generalized causal inference, but is never perfect.

Shadish et al. (2002) recommend assessing causal claims by investigating the operations (persons, settings, treatments, and outcomes) of the study for construct validity and external validity, as graphically illustrated in Figure 3.3. Construct validity and external validity are determined and argued through five guiding principles, which will be defined next.

Principle one, surface similarity, identifies the prototypical characteristics of the target operations (persons, settings, treatments, and outcomes). In other words, important attributes of the causal claim must be defined clearly. Principle two, ruling out irrelevancies, identifies factors that were present in the study but were irrelevant to the study. Well-designed studies include a heterogeneity of irrelevancies so that any potential confounds that are misdiagnosed as irrelevancies do not have an unintended effect on the causal claim. The third principle, making discriminations, identifies which features of the study limit generalization. Interpolation and extrapolation identify how evidence within the causal continuum may be inferred within the range of measurement and beyond the range of measurement. Lastly, the principle of causal explanation seeks to find what parts of the treatment affect what parts of the outcome. It
investigates similar and dissimilar explanations and attempts to show how competing explanations are inferior to the generalized causal inference of the current study. According the Shadish et al. (2002), these principles constitute the ability to, “specify (a) which parts of the treatment (b) affect which parts of the outcome (c) through which causal mediating processes in order to accurately describe the components that need to be transferred to other situations to reproduce the effect” (p. 358). This process provides the roadmap by which generalized causal inference will be argued in Chapter 5.

Limitations

This study was limited by several sampling and methodological weaknesses. Restrictions placed on the eligibility of schools, e.g., schools starting band in fifth grade with willing
instructors, reduced the population validity of the study. This was because the accessible population became a smaller subset of the target population with every restriction. This affected the generalizability of the study but was necessary to feasibly conduct the study within the given budget, time frame, and resources.

Additional sampling issues arose from the willingness of students and parents to participate in the study. Those students who refused or were not able to participate eliminated valuable data from the pool. The experimental band groups were overloaded with extra students to accommodate an anticipated dropout rate of 10%.

Limiting data collection to only one year of study was another limitation. Under ideal circumstances, this study would complete a longitudinal analysis that would measure students’ long-term change. Additionally, this study would be conducted as students enter their second or third year of band study, allowing more time for students to adopt and engage in meaningful literacy techniques. However, ethical considerations (denying instruction to create a control group) and the inability to create a truly randomized sample among second and third-year band students prevented the feasibility of such a study.

The tests themselves were limited in their measurement capabilities. The NWEA MAP reading tests only measure reading skills. Even though reading scores included component subtests, they did not include other areas of literacy as indicated in the literacy model presented in Chapter 2, such as listening, speaking, writing, or understanding contextual knowledge and skills. Similarly, the WFPS only measured reading and performance while neglecting the remaining areas deemed important to the overall literacy model.

Methodologically, quantitative studies are criticized for their inability to truly establish a causal relationship, particularly given the complexity of social behavior (Cohen, Manion, &
Morrison, 2011; Gall, Gall, & Borg, 2007). This study could not control for all extraneous influences; however, it was hoped that the randomized assignments helped mitigate unexplained influences by distributing random error equally across all sample groups.

Conclusion

Chapter 3 presented the methodology that was used to investigate the research questions. The chapter began with an overview of the quantitative research design and its rationale. Next, the chapter explored the sampling procedures, treatment procedures, data collection strategies, and data analysis strategies. The chapter concluded by declaring the limitations of the study. The next chapter will present the data that were collected during this study.
CHAPTER 4

RESULTS

The purpose of this study was to examine the reciprocal effects of music literacy instruction and textual literacy instruction among elementary band students in conventional and literacy-enriched contexts. Chapter 4 will present the results of the data collection and analysis. It will begin with an exploration of the school, teacher, and students who participated in the study. The demographic variables will be examined to gauge their potential influence on the outcome of the main tests. The method for data preparation will follow, which includes the process for data screening and the handling of outliers. The chapter will conclude with an analysis of the assumptions for each test followed by an account of the test results associated with each research question.

Research Questions

The following research questions concern textual literacy achievement:

1. How does conventional band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?

2. How does literacy-enriched band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?
3. Does literacy-enriched band instruction yield a different effect from conventional band instruction on textual literacy scores among elementary students?

The following research question concerns music literacy achievement:

4. Does literacy-enriched band instruction yield a different effect from conventional band instruction on musical literacy scores among elementary students?

Hypotheses

Hypothesis 1: Students participating in a conventional band environment will exhibit greater textual literacy growth than students who do not participate in band when controlling for pretest scores. In other words, \( H_1: (\mu_1 > \mu_0) \) where 0 is the control group (non-band) and 1 is the conventional band group. The null hypothesis states that no difference exists between the groups or the control group outscores the conventional band group, or \( H_0: (\mu_1 \leq \mu_0) \).

Hypothesis 2: Students who participate in a literacy-enriched band environment will exhibit greater textual literacy growth than students who do not participate in band when controlling for pretest scores. This is represented with \( H_2: (\mu_2 > \mu_0) \) where 0 is the control group (non-band) and 2 is the literacy-enriched band group. The null hypothesis states that no difference exists between the groups or the control group outscores the literacy-enriched band group, or \( H_0: (\mu_2 \leq \mu_0) \).

Hypothesis 3: Students in the conventional band group will exhibit different textual literacy scores than students in the literacy-enriched band group when controlling for pretest scores. The null hypothesis suggests that no difference exists between the group means.
Otherwise stated, H₃: (µ₁ ≠ µ₂) where 1 is the conventional band group and 2 is the literacy-enriched band group, while the null hypothesis is similarly coded H₀: (µ₁ = µ₂).

Hypothesis 4: Students in the conventional band group will exhibit different musical literacy scores than students in the literacy-enriched band group. The null hypothesis suggests that no difference exists between the group means. Otherwise stated, H₄: (µ₁ ≠ µ₂) where 1 is the conventional band group and 2 is the literacy-enriched band group, while the null hypothesis is coded H₀: (µ₁ = µ₂).

Sample Participants

The participating school was chosen because it met the criteria for eligible sites. Namely, 1) the school administered the NWEA MAP test at least twice a year as part of its extant assessment framework, 2) the school’s curricular band program began in grade five or later, and 3) the school allowed access for the study, including facilities and staff. The school is an urban elementary school in Illinois that serves grades K through six. It hosts a dual language program throughout all seven grade levels. The average class size is 23 students. According to the school’s Illinois State Report Card (2015), the school comprises a 79.6% Hispanic, 7.5% Caucasian, and 8.9% black population. Regarding SES and native language speakers, 90.8% of students are considered low income and 58.4% are English language learners (English is not their primary language).

The selected school did not represent the population of interest well. According to the National Center for Educational Statistics (NCES, 2017), the ethnic diversity of public school students nationwide consists of 50% White, 25% Hispanic, 16% African American, and 5% Asian. However, the trend is shifting, and the percentage of Hispanics is rapidly rising. The
NCES estimates that by 2026, Whites and African American distributions will lower to 45% and 15%, and Hispanics and Asians will rise to 29% and 6% respectively. Although the chosen school still did not represent the overall demographic consistency of all public school students, it did represent an increasingly greater number of urban schools. Therefore, there was still value in investigating the sample as an accurate representation of a primarily Hispanic, low-SES urban school. The discussion will take this into account.

The band instructor at this school was a state-licensed teacher with five years of teaching experience. She agreed to serve as the instructor for the experimental groups in the study and abide by the terms and protocols of this study. Students benefitted from the consistency of having the same band instructor in grades five through eight since she was the band instructor at their elementary school and their subsequent junior high school. She also helped manage communications with the fourth-grade classroom teachers, custodial staff, and administration.

Some student data were removed from the final analysis. The overall number of participants was diminished by non-study related factors. Two students moved out of the participating school and three students who provided consent and were invited to participate in the study did not attend the instructional sessions. To preserve the integrity of the random group assignments, six students who switched groups due to scheduling conflicts were also not included in the statistical tests. Thus, out of 43 consenting participants, 32 student scores were included in the analysis. This represented a 26% decrease in available data.

The 11 removed scores were analyzed to determine whether the missing data would have potentially changed the outcome of the statistical tests. All 11 pretest scores fell within 2.27 RIT units of the sample mean on the MAP overall reading test, literature subtest, vocabulary subtest, and the informational text subtest (see table 4.1). Thus, they were not significantly different at
the $\alpha = .05$ level. Similarly, no statistically significant differences were found among covariate-adjusted posttest scores. Only nine of the removed scores were available for posttest analysis, which fell within 7.00 RIT units of the sample mean. Therefore, the excluded data were deemed statistically equivalent to the sample and did not pose a threat to alter any conclusions.

Table 4.1

<table>
<thead>
<tr>
<th></th>
<th>Pretest (Covariate)$^a$</th>
<th>Posttest (Dependent Variable)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M^c$ (SD)</td>
<td>$t(20)$</td>
</tr>
<tr>
<td><strong>Overall Reading Test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removed Cases</td>
<td>191.64 (18.76)</td>
<td>.106</td>
</tr>
<tr>
<td>Control</td>
<td>192.36 (12.78)</td>
<td></td>
</tr>
<tr>
<td><strong>Literature Subtest</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removed Cases</td>
<td>191.73 (19.81)</td>
<td>.312</td>
</tr>
<tr>
<td>Control</td>
<td>194.00 (13.81)</td>
<td></td>
</tr>
<tr>
<td><strong>Vocabulary Subtest</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removed Cases</td>
<td>190.64 (19.53)</td>
<td>.087</td>
</tr>
<tr>
<td>Control</td>
<td>191.27 (14.33)</td>
<td></td>
</tr>
<tr>
<td><strong>Informational Text Subtest</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removed Cases</td>
<td>192.64 (18.99)</td>
<td>-.079</td>
</tr>
<tr>
<td>Control</td>
<td>192.09 (12.91)</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** $M =$ Mean.  SD = Standard Deviation.  SEM = Standard Error of the Mean.

$^a$ Removed Cases $n = 11$ and Control $n = 11$.  $^b$ Removed Cases $n = 9$ and Control $n = 11$.  $^c$ Values are reported as NWEA RIT units.
The 32 remaining participants consisted of 18 females (56%) and 14 males. Five students were ten years of age (16%), while the remaining students were nine. Two students took private music lessons outside of school, but for less than one year. One student studied ukulele and the other student studied piano. Participants roughly reflected the ethnic distribution of the school, although the sample had more Asian students and less Caucasian students than the school average (Figure 4.1).

![Pie chart showing ethnicity distribution of study participants.](image)

**Figure 4.1.** Ethnicity of study participants.

Similarly, the socio-economic distribution of the participants roughly reflected the overall school (Illinois State Report Card, 2015). Parents’ status under the Free and Reduced Lunch Act served as a proxy for socio-economic standing. Sixty-nine percent of participants received a free lunch, and 19% received a reduced rate lunch, which equaled a total of 88% receiving assistance under the National School Lunch Program (Figure 4.2). This compared to a nationwide average
of 45% participating in the program among African Americans and Hispanics in high-poverty schools (NCES, 2016). Therefore, the selected school was found to have a higher rate of low SES students than the national average.

Family mobility rate and parental involvement have been found to have an impact on the academic success of music students by previous research (Costa-Giomi, 2012; Kinney, 2008). The mobility rate for this study was relatively low in this study sample. Only four students (13%) had lived at their present address for less than two years. Data were also collected investigating the presence of one or two parents in the child’s home. Four students (13%) lived in one parent households, 20 students (63%) lived in two-parent households, and the remaining six students (19%) preferred not to indicate their parental involvement status.

Students were selected from all five fourth-grade classroom teachers—Brown, Collins, Garrison, Green, and Stanton—to maintain internal validity. The classroom teachers are
represented by a pseudonym in this study to preserve confidentiality. Figure 4.3 indicates the
distribution of students by teacher.

Also, to further verify internal validity, the three groups (control, conventional, and
literacy-enriched) were investigated for pretest differences among the entire class of fourth-
graders using a one-way ANCOVA ($N = 104$). Although this procedure was not part of the
primary analysis, it helped affirm that the sample groups were representative of the entire fourth-
grade class. Since MAP data were divided by the overall scores and three subtest scores,
separate ANCOVA procedures were administered for each score set. No significant differences
were found among the class mean, control group mean, or experimental group means on any of
the four reading pretests at a criteria of $\alpha = .05$ (Appendix H). Additionally, no students in the
sample required an Individualized Education Program (IEP) or a Section 504 plan from the Rehabilitation Act of 1973 (see Stanberry, 2017).

Due to the random assignment of students to the control and experimental groups, any confounding effects of these demographic data were dispersed evenly throughout the groups, minimizing their effects. However, despite randomization, group means will sometimes vary probabilistically, introducing confounding bias to the scores, especially among smaller sample sizes (Cook & Sinha, 2006). Although not part of the primary analysis, several $t$ tests were performed to find if any *a priori* influential relationships existed between the dichotomized demographic variables (gender, ethnicity, and SES) and their pretest and posttest score differences (gain scores). The tests examined the overall MAP reading RIT score in addition to the three reading subtests. After testing for homogeneity, the $t$ tests revealed that there were no significant relationships between the demographic groups and their corresponding gain score differences (Table 4.2).

In addition to examining the influence of gender, ethnicity, and SES using $t$ tests, the influence of the general classroom teacher was measured using a one-way ANOVA. This examination helped eliminate the potential bias of classroom teacher instruction as an influence on scores. An ANOVA was used because the teacher variable could not be dichotomized since there were four classroom teachers, and an ANOVA is appropriate when examining a single factor with more than two categories. After affirming the test assumptions, the ANOVA showed no significant relationships between the classroom teacher and the gain scores (Table 4.3).
Table 4.2
Gender, Ethnicity, and SES as Potential Confounds on Pretest and Posttest Differences

<table>
<thead>
<tr>
<th></th>
<th>Dichotomized Values (n)</th>
<th>95% CI</th>
<th>t(30)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MD</td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>Overall Reading Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Female (18), Male (14)</td>
<td>.929</td>
<td>-6.54</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Hispanic (25),</td>
<td>.669</td>
<td>-6.08</td>
<td>7.41</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic (7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>Free (22), Non-Free (10)</td>
<td>4.37</td>
<td>-10.17</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Literature Subtest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Female (18), Male (14)</td>
<td>.373</td>
<td>-7.08</td>
<td>6.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Hispanic (25),</td>
<td>.554</td>
<td>-12.46</td>
<td>11.36</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic (7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>Free (22), Non-Free (10)</td>
<td>6.57</td>
<td>-13.32</td>
<td>.172</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary Subtest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Female (18), Male (14)</td>
<td>1.38</td>
<td>-9.90</td>
<td>7.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Hispanic (25),</td>
<td>4.38</td>
<td>-5.74</td>
<td>14.49</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic (7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>Free (22), Non-Free (10)</td>
<td>1.12</td>
<td>-8.02</td>
<td>10.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informational Text Subtest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Female (18), Male (14)</td>
<td>4.59</td>
<td>-12.53</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Hispanic (25),</td>
<td>2.27</td>
<td>-11.99</td>
<td>7.44</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic (7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES</td>
<td>Free (22), Non-Free (10)</td>
<td>4.60</td>
<td>-13.13</td>
<td>3.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: SES = Socioeconomic status as determined by free lunch, reduced-rate lunch, or no lunch subsidies awarded. Free = Students receiving a free lunch, Non-Free = Students receiving a reduced rate lunch, full price lunch, or did not indicate lunch status. MD = Mean difference between pretest and posttest scores. CI = Confidence interval. LL = Lower limit; UL = Upper limit.
Table 4.3
Classroom Teacher as a Potential Confound on Reading Test Differences

<table>
<thead>
<tr>
<th>Reading Test Type</th>
<th>MD</th>
<th>SD</th>
<th>F(4, 27)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Reading Test</td>
<td>1.16</td>
<td>6.85</td>
<td>2.034</td>
<td>.118</td>
</tr>
<tr>
<td>Literature Subtest</td>
<td>1.72</td>
<td>9.06</td>
<td>1.237</td>
<td>.319</td>
</tr>
<tr>
<td>Vocabulary Subtest</td>
<td>0.44</td>
<td>11.54</td>
<td>0.917</td>
<td>.468</td>
</tr>
<tr>
<td>Informational Text Subtest</td>
<td>1.00</td>
<td>11.49</td>
<td>1.009</td>
<td>.420</td>
</tr>
</tbody>
</table>

Note: MD = Mean difference between the pretest and posttest between all teachers. SD = Standard deviation. *N = 32.*

Results for Research Questions 1, 2, and 3

Research questions 1, 2, and 3 were all investigated using planned contrasts from the same ANCOVA procedure. Research question 1 compared group 0 (control) to 1 (conventional), research question 2 compared group 0 (control) to 2 (literacy-enriched), and research question 3 compared group 1 (conventional) to 2 (literacy-enriched). Since the reading subtest scores were available in addition to the overall reading scores, the subtest scores were also included in the analysis. This was accomplished by using an ANCOVA procedure for each reading subtest in addition to the ANCOVA procedure for the overall reading test. Each procedure consisted of investigating assumptions, completing the omnibus test, then completing the three planned contrasts to address each research question.

An *a priori* decision was made to increase the alpha rate to .10 if the sample size fell below the threshold of 72 participants to maintain a minimal level of statistical power. Since the
actual sample size was 32, an alpha rate of .10 was established for the ANCOVA procedure. Under these conditions, an effect of \( d = 1.0 \) would produce a power of .78, and an effect of \( d = .75 \) would produce a power of .54. An effect of either magnitude would be difficult but achievable given similar literature (Butzlaff, 2000).

If each hypothesis corresponded with a single planned contrast, multiplicity would not have been a concern. However, the addition of the three subtest planned contrasts created a total of four comparisons for each hypothesis. To preserve the family-wise error rate at .10, a Bonferroni correction was applied to the \( p \) statistic of the contrasts using the formula \( p < \alpha / k \), where \( k \) was the number of comparisons, creating a critical \( p \)-value of .025. Bonferroni was chosen over Tukey or Dunnett because the number of comparisons was relatively small and the Bonferroni adjustment was not overly conservative (Seltman, 2013).

**ANCOVA Assumptions**

Prior to the administration of the omnibus test and the three planned contrasts, ten assumptions were investigated to observe whether the data were appropriate for an ANCOVA model (Field, 2013; Laerd Statistics, 2015). These assumptions were applied to the overall reading scores followed by the three reading subtest scores. The first four assumptions were related to the design of the study. The assumptions suggested that the study contained 1) a single, continuous dependent variable (posttest scores), 2) a single, categorical factor (group assignment), 3) a single, continuous covariate variable (pretest scores), and 4) participants that did not belong to more than one group (independence of observations). The design of the study ensured that these assumptions were observed for all four ANCOVA procedures (overall reading
scores, literature subtest scores, vocabulary acquisition and use subtest scores, and informational text subtest scores).

Overall Reading Test Assumptions

Six additional assumptions regarding the viability of the data for an ANCOVA were tested statistically, starting with the overall reading test scores. First, the covariate (pretest) was linearly related to the dependent variable (posttest) within all three groups (control, conventional, and literacy-enriched). This was tested by visually inspecting a scatter plot using regression lines to represent the linear approximation of each group (Figure 4.4). The plot showed that the dependent variable and covariate had a linear relationship (i.e., a straight line) between all three groups of the factor.

The second assumption required homogeneity of regression slopes. In other words, there was to be no interaction between the covariate (pretest) and the factor (group), which is typically indicated by parallel regression lines on the scatter plot. Unfortunately, Figure 4.4 showed intersecting regression lines between the control group (group 0) and the conventional group (group 1), thus violating the assumption. To verify the violation, the degree of homogeneity was tested using an ANCOVA with the covariate and group interaction term selected as a custom model. This test revealed that the interaction term was not significant, $F(2, 26) = .070$, $p = .933$. Therefore, the assumption was determined not to be violated.

The third assumption required that the dependent variable standardized residuals were approximately normally distributed for all three groups of the factor. Standardized residual values were created and stored when the omnibus ANCOVA was administered. Using those
standardized residuals, the assumption of normal distribution was met by both visually
inspecting Q-Q plots and interpreting Shapiro-Wilk's test ($p > .05$). The fourth assumption stated
that homoscedasticity of error variances existed within each group. This assumption was tested
by visually inspecting a scatter plot between standardized residuals against predicted values. The
plot indicated no patterns or uneven spread. Therefore the assumption was met (Figure 4.5).
The fifth assumption required homogeneity of variances, which meant that the variance of residuals was roughly equal between all three groups. This assumption was investigated and met as indicated by Levene’s Test of Homogeneity of Variances ($p = .176$). The sixth assumption required that there were no standardized residual outliers greater than ±3 standard deviations. The standardized residuals were sorted and inspected, resulting in no cases at or beyond ±3 standard deviations, affirming the assumption of no outliers. Therefore, all assumptions of fit for the overall reading test data to the ANCOVA model were satisfied.
**Literature Subtest Assumptions**

The six assumptions that were statistically tested for the overall reading test scores were also applied to the literature subtest. For assumption 1, a scatter plot revealed a linear relationship between the dependent variable and the covariate. For assumption 2, the homogeneity of regression slopes showed that the interaction term between the group and covariate were not significant, $F(2, 26) = .186, p = .831$. For assumption 3, a Shapiro-Wilk test showed that standardized residuals were normally distributed ($p > .05$). For assumption 4, homoscedasticity was found by visually inspecting standardized residuals plotted against predicted values. Assumption 5 investigated the homogeneity of variances, which was indicated by Levene’s Test ($p = .075$). The final assumption (6) cautioned against outliers greater than ±3 standard deviations. An investigation of sorted standardized residuals revealed no outliers. Therefore, all assumptions were met, indicating that the literature subtest data were a fit for the ANCOVA model.

**Vocabulary Acquisition and Use Subtest Assumptions**

The six assumptions were then applied to the vocabulary acquisition and use subtest scores. The linear relationship between the dependent variable and covariate was visually inspected and affirmed. The interaction term between group and covariate was not significant, indicating that the assumption of homogeneity of regression slopes was present, $F(2, 26) = 1.154, p = .331$. Standardized residuals were normally distributed according to the Shapiro-Wilk test ($p > .05$). A plot containing standardized residuals and predicted values showed homoscedasticity. Levene’s Test showed that the variances were homogeneous ($p = .460$), and
the standardized residuals revealed no outliers beyond ±3 standard deviations. Therefore, the data were deemed to be a good fit for the ANCOVA procedure.

**Informational Text Subtest Assumptions**

The same six assumptions were finally applied to the informational text subtest. A linear relationship between the dependent variable and covariate was present, as was homogeneity of regression slopes $F(2, 26) = .487, p = .620$. The Shapiro-Wilk test once again showed normally distributed standardized residuals ($p > .05$). Homoscedasticity was indicated by inspecting a plot of standardized residuals and predicted values. Levene’s Test suggested that the homogeneity of variances was present ($p = .366$). Lastly, standardized residuals contained no outliers beyond ±3 standard deviations. These data were also deemed appropriate for an ANCOVA procedure.

**Omnibus ANOVA Results for the MAP Reading Test and Subtests**

The omnibus ANCOVA was administered to determine the degree to which the group assignment affected the overall reading posttest scores. The group means were adjusted by the covariate then compared, resulting in differences among all three groups (Table 4.4). A Bonferroni correction was not applied to the omnibus tests because they were not exploratory or reiterative and thus presented no threat for a Type I error. The alpha rate was therefore retained at .10. However, since the hypothesis indicated directionality, a one-tailed test was used.

The omnibus ANCOVA rejected the null hypothesis $F(2, 28) = 2.11, p = .070$ (one-tailed). Practically speaking, this indicated that at least one of the groups was significantly different from another group, but the omnibus ANCOVA did not specify which group was different. Planned contrasts were used to identify those groups and will be discussed in the next
Table 4.4

Overall MAP Pretest and Posttest Reading Scores

<table>
<thead>
<tr>
<th></th>
<th>Pretest (Covariate)</th>
<th>Posttest (Dependent Variable)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>$M$</td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>192.36</td>
</tr>
<tr>
<td>Group 1 (Conventional)</td>
<td>11</td>
<td>199.36</td>
</tr>
<tr>
<td>Group 2 (Literacy-enriched)</td>
<td>10</td>
<td>189.80</td>
</tr>
</tbody>
</table>

Note: MAP refers to the NWEA Measures of Academic Performance Test; adj. $M$ = Mean scores are adjusted by the covariate; $SEM$ = Standard error of the mean.

section. Effect sizes for the omnibus ANCOVA are not particularly informative (Field, 2013) and were therefore withheld in favor of the effect size calculations for the planned contrasts. Assuming that the treatment actually had no effect, the observed differences (or greater) would have occurred approximately 7% of the time by chance if this study was repeated many times. Probabilistically, this finding suggests that the treatment was correlated with the outcome.

Next, the same omnibus ANCOVA procedure was administered to the three reading subtests. The planned contrasts, not the omnibus test, contained the results of interest to this study, but the omnibus results are presented for the sake of completeness. The subtest posttest scores were adjusted by the covariate, and all subtests revealed practical differences within and between each group (Table 4.5). The omnibus ANCOVA on the literature subtest data rejected
Table 4.5
MAP Reading Subtest Pretest and Posttest Scores

<table>
<thead>
<tr>
<th>Subtest/Group</th>
<th>n</th>
<th>Pretest (Covariate)</th>
<th>Posttest (Dependent Variable)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( M )</td>
<td>( SD )</td>
</tr>
<tr>
<td><strong>Literature</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>194.00</td>
<td>13.81</td>
</tr>
<tr>
<td>Group 1 (Conventional)</td>
<td>11</td>
<td>196.55</td>
<td>9.16</td>
</tr>
<tr>
<td>Group 2 (Literacy-enriched)</td>
<td>10</td>
<td>190.50</td>
<td>15.74</td>
</tr>
<tr>
<td><strong>Vocabulary</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>191.27</td>
<td>14.33</td>
</tr>
<tr>
<td>Group 1 (Conventional)</td>
<td>11</td>
<td>199.00</td>
<td>6.59</td>
</tr>
<tr>
<td>Group 2 (Literacy-enriched)</td>
<td>10</td>
<td>188.70</td>
<td>12.53</td>
</tr>
<tr>
<td><strong>Informational Text</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>11</td>
<td>192.09</td>
<td>12.91</td>
</tr>
<tr>
<td>Group 1 (Conventional)</td>
<td>11</td>
<td>200.45</td>
<td>7.48</td>
</tr>
<tr>
<td>Group 2 (Literacy-enriched)</td>
<td>10</td>
<td>190.20</td>
<td>11.78</td>
</tr>
</tbody>
</table>

*Note: MAP refers to the NWEA Measures of Academic Performance Test; \( adj. M \) = Mean scores are adjusted by the covariate; \( SEM \) = Standard error of the mean.*
the null hypothesis $F(2, 28) = 3.353, p = .025$ (one-tailed). Similarly, the omnibus ANCOVA on
the informational text subtest data also rejected the null hypothesis $F(2, 28) = 2.332, p = .058$
(one-tailed). On the contrary, the omnibus ANCOVA on the vocabulary subtest data retained the
null hypothesis $F(2, 28) = .527, p = .298$ (one-tailed). Observed power for the omnibus
ANCOVA on the vocabulary subtest was .21, well below the desired power of .80. The lack of
power was due to a combination of unexplained score variability and low sample size, indicating
that this model did not fit the data well. Since the null hypothesis was affirmed, there was no
need to perform planned contrasts on the vocabulary subtest. However, the research questions
necessitated the need to calculate the planned contrasts despite the omnibus results.

**Research Question 1: Control Group Versus the Conventional Group**

The planned contrasts provided specific information about the relationship among each of
the three groups. However, SPSS was not capable of performing planned contrasts within an
ANCOVA procedure. To conduct the planned contrasts, the standard ANCOVA setup process
was modified by selecting “Compare Main Effects” using an LSD (none) confidence level
adjustment. This provided covariate-adjusted posttest scores that were not affected by post-hoc
adjustments. Additionally, SPSS assumed a two-tail alpha split; to convert the formula to a one-
tail calculation, the alpha level was doubled. Therefore, the significance level input field was
adjusted to .20 to produce a one-tailed alpha level of .10. This adjustment produced 90%
confidence intervals (CI) based on the covariate-adjusted dependent variable. Only the lower
level CI was reported since the alpha level was pooled on the right side of the model. Finally,
SPSS automatically provided two-tailed $p$ values, which were divided by 2 to obtain the correct
one-tailed values. The results of the contrasts were verified by completing $t$ tests on the adjusted
means, entering the adjusted means into an online statistical calculator (Mathercracker, 2017), and consulting a table of critical $t$ values.

Research questions one, two, and three utilized four planned contrasts each. Therefore, the planned contrasts required a Bonferroni correction to avoid inflating the experiment-wise error rate beyond $a = .10$. This altered the critical $p$-value for each test from .10 to .025. The contrasts were orthogonal, meaning all comparisons were conducted against separate and independent hypotheses and groups were not compared twice in the same analysis.

The planned contrasts comparing the control group to group 1 (the conventional group) showed no statistically significant increases after applying the Bonferroni adjustment (Table 4.6). Descriptively, the data did indicate that conventional band students performed well on the

<table>
<thead>
<tr>
<th>Test/Subtest</th>
<th>Difference (Group 1- Group 0)</th>
<th>$MD$</th>
<th>$SE$</th>
<th>$90%$ CI (LL)$^a$</th>
<th>$t(20)$</th>
<th>$p^b$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Reading</td>
<td>-0.38</td>
<td>2.97</td>
<td>-5.42</td>
<td>-0.126</td>
<td>.450</td>
<td>.123</td>
<td></td>
</tr>
<tr>
<td>Literature</td>
<td>6.61</td>
<td>3.62</td>
<td>0.36</td>
<td>1.823</td>
<td>.042</td>
<td>.693</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>1.17</td>
<td>5.13</td>
<td>-7.55</td>
<td>0.231</td>
<td>.410</td>
<td>.146</td>
<td></td>
</tr>
<tr>
<td>Informational Text</td>
<td>-7.68</td>
<td>3.90</td>
<td>-14.27</td>
<td>-2.011</td>
<td>.971</td>
<td>.754</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6

MAP Score Differences Between the Control Group and Group 1 (Conventional Instruction)

$Note$: MAP refers to the NWEA Measures of Academic Performance Test. $MD =$ Mean score difference between students in group 1 and group 0. $SE =$ Standard error.

$^a$ Only lower limit reported for a one-tailed test. Values above 0 indicate a potentially meaningful difference. $^b p$-values were significant at the .025 level due to a Bonferroni adjustment, which comprised an experiment-wise alpha of .10.
literacy subtest as indicated by a confidence interval that did not span zero and a \( p \)-value of .042. However, since the critical value of \( p \) was adjusted to .025 to avoid a type I error, the model did not possess the power necessary to indicate a statistically significant difference between the control group and the conventional band group.

Statistical power was particularly weak due to the small sample size and large unexplained variance. None of the tests reached the intended target power of .80, meaning that the likelihood of committing a type II error (the beta level) was greater than 20%. Under those circumstances, it was possible that the model failed to identify a relationship that was true in the population. Therefore, the lack of significance should be interpreted with caution. The premise of the first alternative hypothesis was not proven false; rather, it was determined to be inconclusive based on the current sample and statistical model. Additionally, effect sizes were not calculated due to the lack of statistical significance. Therefore, nothing may be inferred from the model statistically, as the model retained the null hypothesis claiming equal means between the control group and group 1.

**Research Question 2: Control Group Versus the Literacy-Enriched Group**

A one-tailed \( p \)-value was used to determine statistical significance since the research question asked how literacy-enriched band instruction increased reading scores compared to non-band students. Like the previous planned contrast, decreasing scores were not anticipated in accordance with previous music research (Barlar, 2010; Neuharth, 2000; Wachtel, 2006). Descriptively, the planned contrasts comparing the control group to group 2 (the literacy-enriched group) did show a large increase between the control group and the literacy-enriched group in the overall reading scores and literature subtest scores. These increases were indicated
by confidence intervals that failed to span zero. However, one test produced a statistically 
significant improvement in test scores between the control group and the literacy-enriched 
group—the literature subtest (Table 4.7).

Table 4.7

MAP Score Differences Between the Control Group and Group 2
(Literacy-Enriched Instruction)

<table>
<thead>
<tr>
<th>Test/Subtest</th>
<th>Difference (Group 2 - Group 0)</th>
<th>$M$</th>
<th>$SE$</th>
<th>90% CI (LL)$^a$</th>
<th>$t$(19)</th>
<th>$p^b$</th>
<th>$d$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Reading</td>
<td></td>
<td>5.25</td>
<td>2.94</td>
<td>0.18</td>
<td>1.768</td>
<td>.047</td>
<td>-</td>
<td>.673</td>
</tr>
<tr>
<td>Literature</td>
<td></td>
<td>9.22</td>
<td>3.72</td>
<td>2.80</td>
<td>2.484</td>
<td>.011**</td>
<td>1.08</td>
<td>.874</td>
</tr>
<tr>
<td>Vocabulary</td>
<td></td>
<td>5.07</td>
<td>5.08</td>
<td>-3.80</td>
<td>0.987</td>
<td>.168</td>
<td>-</td>
<td>.376</td>
</tr>
<tr>
<td>Informational Text</td>
<td></td>
<td>-0.28</td>
<td>3.80</td>
<td>-6.93</td>
<td>-0.070</td>
<td>.528</td>
<td>-</td>
<td>.113</td>
</tr>
</tbody>
</table>

Note: MAP refers to the NWEA Measures of Academic Performance Test. $M$ = Mean scores among students in the group. $SE$ = Standard error.

$^a$ Only lower limit reported for a one-tailed test. Values above 0 indicate a potentially meaningful difference. $^b$ $p$-values are significant at the .025 level due to a Bonferroni adjustment, which comprised an experiment-wise alpha of .10.

** $p < .025$.

Statistical power was improved due to larger differences between the group means, but 
except for the literature subtest, the model still failed to meet the .80 target. The model required 
a larger sample and less unexplained variance among the test scores. Thus, the second 
alternative hypothesis remained inconclusive for the overall reading test, vocabulary subtest, and 
the informational text subtest. On the other hand, the literature subtest scores were statistically 
significantly higher between the literacy-enriched group and the conventional group, producing a
power level of .874. This is the only test that met the .80 power threshold. Since the evidence indicated a statistically significant relationship, the next step was to determine the size of the effect.

Two types of measures are typically used to evaluate effect size, and they are often categorized by designs that express the mean difference between two groups in units of standard deviation, such as Cohen’s $d$ and Hedges’ $g$, or designs that express the proportion of variance in the dependent variable accounted for by the independent variable, such as $r$ squared, eta squared, and omega squared (Grace-Martin, 2011). Cohen’s $d$ was calculated using the formula $d = (M_1 - M_2) / SD_{pooled}$, which produced a large observed effect where there was over a one standard deviation difference between the control group and group 2 ($d = 1.08$). The effect was extremely large in the context of existing music literacy research (Standley, 2008). Practically, this may be interpreted to mean that students who were taught band using a literacy-enriched paradigm scored an average of 7 points higher on the MAP literature subtest than students who did not participate in band.

Eta squared is the proportion of total variance that is attributed to an effect, and is calculated using the formula $\eta^2 = SS_{effect} / SS_{total}$ (Field, 2013). The literature subtest model produced three sources of variance—the covariate, which accounted for 67% of the variance ($\eta^2 = .671$), the group assignment, which accounted for 19% of the variance ($\eta^2 = .193$), and residual error, which accounted for the remaining 14% of the variance. According to Cohen (1988), Miles and Shevlin (2001), and related music literacy research (Butzlaff, 2000; Winner & Cooper, 2000), an eta squared value of .193 (explained by the group assignment) represents a large proportion of variance explained in a one-way ANOVA model.
Research Question 3: Conventional Group Versus the Literacy-Enriched Group

A two-tailed test was used to investigate the planned contrasts between group 1 (the conventional group) and group 2 (the literacy-enriched group), since, unlike research questions 1 and 2, no directionality was implied in research question 3. The Bonferroni correction again called for a critical $p$-value of .025 to avoid inflated alpha levels due to multiplicity. As a result, none of the planned contrasts produced statistically significant values (Table 4.8). Thus, the planned contrasts failed to reject the null hypothesis that the means were significantly different between group 1 and group 2.

Table 4.8

MAP Score Differences Between the Group 1 (Conventional Instruction) and Group 2 (Literacy-Enriched Instruction)

<table>
<thead>
<tr>
<th>Test/Subtest</th>
<th>Difference (Group 2 - Group 1)</th>
<th>$M$</th>
<th>$SE$</th>
<th>90% CI (LL, UL)</th>
<th>$t$(19)</th>
<th>$p^a$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Reading</td>
<td></td>
<td>5.62</td>
<td>3.13</td>
<td>(-10.98, -0.26)</td>
<td>1.813</td>
<td>.086</td>
<td>.558</td>
</tr>
<tr>
<td>Literature</td>
<td></td>
<td>2.62</td>
<td>3.76</td>
<td>(-9.07, 3.83)</td>
<td>0.703</td>
<td>.491</td>
<td>.181</td>
</tr>
<tr>
<td>Vocabulary</td>
<td></td>
<td>3.90</td>
<td>5.40</td>
<td>(-12.95, 5.17)</td>
<td>0.743</td>
<td>.467</td>
<td>.186</td>
</tr>
<tr>
<td>Informational Text</td>
<td></td>
<td>7.40</td>
<td>4.08</td>
<td>(-14.23, -0.59)</td>
<td>1.879</td>
<td>.076</td>
<td>.567</td>
</tr>
</tbody>
</table>

Note: MAP refers to the NWEA Measures of Academic Performance Test. $M$ = Mean scores among students in the group. $SE$ = Standard error.

$^a$ $p$-values are significant at the .025 level due to a Bonferroni adjustment, which comprised an experiment-wise alpha of .10.
Descriptively, the overall reading test and informational text subtest produced 90% confidence intervals that did not span zero for the overall reading test and the informational text subtest. Typically, that would indicate statistical significance. However, the Bonferroni adjustment established a much more conservative criterion for statistical significance, causing both $p$ values to fail to reach the critical level. Statistically, the models were not powerful enough to meet the .80 threshold for power or the .025 threshold for $p$, meaning that the hypothesis was deemed inconclusive, warranting the need for further study.

Results for Research Question 4

The Watkins-Farnum Performance Scale (WFPS) was used to test the differences of musical reading and playing abilities between Group 1 (conventional instruction) and Group 2 (literacy-enriched instruction) after 17 weeks of instruction. The playing test was administered by the instructor during the week before the students’ final concert performance. All tests were audio recorded using an iPad.

The test was proctored consistently for each student. Students received the following verbal instructions as indicated on the WFPS administration protocol, “In this test you are to read each exercise exactly as written. Be sure that you hold each note its correct value and observe all markings and signs. The grading system will be quite strict so do your best. The first exercise is to be played at this speed” (Watkins & Farnum, 1954). A metronome was set at 88 bpm and then turned off after the first note was played. The test was scored by counting the number of measures played correctly. Students earned a low score of 0 and a high of 27, $M = 9, SD = 6.09$. Scores were tallied on a copy of the music in real time and then rescored by a second judge using the recordings. The second judge’s scores were compared with the original scores to determine
the inter-rater reliability of the judges and the scoring system. Both were determined to be reliable as indicated by Cronbach’s alpha (.953).

The assumptions of an independent samples $t$ test were examined next. The study design ensured that the first three assumptions were met: 1) the dependent variable contained continuous data, 2) the factor contained categorical data in two groups, and 3) the observations were independent of each other. The fourth assumption required no significant residual outliers. An inspection of a boxplot did reveal one outlier in Group 2 at slightly more than 2 standard deviations from the mean. However, it was determined to be a genuinely unusual yet valid value, not a calculation error, entry error, or measurement error. It therefore required no modification or rejection and was retained in the analysis. The fifth assumption assessed for the normal distribution of scores across all levels of the independent variable. WFPS scores for each group were normally distributed as determined by Shapiro-Wilk's test ($p > .05$). The last assumption assessed and affirmed the homogeneity of variances, as indicated by Levene’s test for equality of variances ($p = .739$).

There were 11 participants in Group 1 (conventional instruction) and 10 participants in group 2 (literacy-enriched instruction). The conventional group scores ($M = 9.91, SD = 5.19$) were slightly higher than the literacy-enriched scores ($M = 9.39, SD = 7.23$). However, the differences were not statistically significantly different, $MD = .522, SE = 2.73, t(19) = .192, p = .850$ (two-tailed), 90% CI [-4.19, 5.23]. This resulted in an extremely low statistical power of .054. The lack of power indicated that this model was a poor fit for the data. As a result, the test affirmed the null hypothesis that the group means were equal.
Summary

This chapter presented an analysis of the text literacy and music literacy data collected for this study. A summary of the research questions and statistical models is provided in Figure 4.6. The boxes that contain the research questions show the $t$ statistics and $p$ values as appropriate for each statistical process. Lower power affected most statistical tests due to the small sample size and unexplained variance, illustrating the need for further investigation with more participants and an improved statistical model.

Although many of the tests affirmed the null hypothesis, the literacy-enriched band group scored significantly higher than the control group on the MAP literature subtest, thereby rejecting the null hypothesis. The effect size for this finding was calculated using both Cohen’s $d$ and eta squared, and was found to be extremely large in the context of educational literature. The next chapter will explore the practical implications of these findings. Additionally, it will explore the factors contributing to the reliability and validity of the results, the limitations of the study, and recommendations for future research.
Note: Contrasts were planned * a priori.  * Value is statistically significant at p < .025.

Figure 4.6. Summary of ANCOVA and independent t test results and their relationship to the four research questions.
CHAPTER 5
DISCUSSION

The field of music does not easily lend itself to positivist or post-positivist investigation. Many have attributed the subjective and artistic nature of music to a lack of empirical study, and thus, a depreciation of music against more structured academic disciplines (see Klieve, Hargreaves, & Morris, 2014). However, this presents a false narrative. Bloom, Engelhart, Furst, Hill, and Krathwohl’s (1956) taxonomy reminds teachers of all grades and all disciplines that the ability to analyze, evaluate, and create generates long-term understanding. In this sense, music is no different from math, science, history, or reading. Yet the idea that music does not lend itself to the same rigor or the same possibility for systematic inquiry may contribute to a perceived indifference surrounding music’s inclusion as a core subject (McMurrer, 2008; Popham, 2001; von Zastrow & Janc, 2004).

The current study provides an opportunity to look critically and empirically at how band teachers may deliver more effective instruction and how that delivery runs parallel to concurrent instruction in ELA. This opportunity is offered not only to help students become better readers in the isolation of an ELA classroom, but also to help them become better musicians in an overwhelmingly language-based beginning band environment. It also allows band directors to compare the efficacy of traditional instructional strategies to more progressive, student-centered strategies. For example, is teacher-centered direction as effective as systematic student-initiated exploration? Could trial-and-error or repetition strategies become more efficient with the
added use of annotating, coding, turn-and-talks, predictions, and summarizations? Alternatively, do student-centered, literacy-enriched strategies require too much time to be effective? The underlying premise of Bruner’s (1977) cognitive structuralist argument is that nothing is learned in isolation. Ideas are regularly bandied about from subject to subject, from day to day. When information can be attached to previous knowledge, long-term memory and understanding are created. Music and ELA are not mutually exclusive. Moreover, although the traditional view of literacy focuses on reading and writing (Harris & Hodges, 1995), a more comprehensive, holistic perspective reveals many more rich interactions between all disciplines and all divisions within the literacy model presented in Chapter 2.

Four themes emerged from the review of literature in Chapter 2. First, cognitive structuralism provides a plausible explanation for the reciprocal growth of music literacy and textual literacy. Second, most global societies sustain or support an unbalanced model of music literacy which includes both background knowledge and communication skills. Third, research has provided little causational evidence relating music literacy to textual literacy, and no causational evidence relating the study of beginning band students and textual literacy, and fourth, a wide body of research indicates a correlation exists between music literacy and textual literacy through quasi-experimental, causal comparative, correlational, and qualitative designs. These themes will be relevant throughout this discussion.

Overall, this study produced mixed results pertaining to the alternative hypotheses proposed in Chapter 1. More evidence was required to make a generalized causal link between conventional band instruction, textual reading growth, and musical reading growth. Despite the lack of statistical significance, however, there was practical utility in the analysis and application
of the theoretical framework, research design, and methodology, which will be explored further in this chapter.

Unlike the conventional band group, the data produced by the literacy-enriched band group suggested that literacy-enriched band instruction may have partially caused reading growth on the MAP literature subtest. The experimental design of the study provided the conditions necessary to defend a causal relationship using Shadish et al.’s (2002) method for investigating generalized causal claims. An analysis and discussion of these claims will also be explored further in this chapter.

Research Question 1: Interpretation and Discussion

Question 1 asked, “How does conventional band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?” The assumption that conventional band instruction and text reading were highly correlated was affirmed in the literature in Chapter 2 (Barlar, 2010, Butzlaff, 2000; Neuharth, 2000; Wachtel, 2006; Winner & Cooper, 2000). Although the current study offered no statistical evidence to show how conventional band instruction caused higher textual reading scores, the study did explore a theoretical explanation for textual literacy growth in addition to several research-based strategies for applying the theory in practice. This information is particularly useful for researchers seeking to replicate the current design or for band teachers seeking to expand their pedagogical skills.

Cognitive structuralism provided the theoretical explanation for the growth of literacy scores among conventional band students by describing how knowledge is created and retained. The physiological processes of reading and interpreting music, combined with the observable
physical responses (buzzing, fingering, breathing, and articulating) as described in Chapter 2, provided the derivative subsumption that allowed cognitive structures to be built and strengthened over time. Most students are capable of receiving or absorbing new information, but they rarely retain the information in their long-term memory unless they link it to existing knowledge, as described by Ausubel (2010). Conventional band instruction provided many opportunities for students to attach new knowledge to existing cognitive structures.

Since research question 1 focused on the link between conventional band instruction and textual literacy, a discussion of how the current study interpreted and implemented parallel processes between conventional band and ELA was prudent. As students progressed sequentially through the Measures of Success text (Sheldon et al., 2010), the instructor used a variety of formative feedback strategies that helped strengthen students’ literacy skills. In particular, the band instructor consistently reinforced, affirmed, and corrected students’ interpretation of both text and music notation throughout every small group lesson and rehearsal. At each meeting, students demonstrated understanding through observable physical responses. Whereas students showed understanding or comprehension in ELA through discussion, written responses, or recitation, students in band showed understanding through performance, fingering, or verbalizing note names and rhythms. Notably, incorrect physical responses in band typically yielded immediate formative feedback, which allowed for a quick reinforcement of fluency and comprehension skills. For example, a trumpet player fingering a “D” 1-2 instead of 1-3 not only looked dissimilar to her peers, but also sounded dissimilar to the point where immediate teacher-, peer-, and self-feedback was possible. In this case, prompt feedback (due to physical cues) reinforced the correct interpretation of written symbols a way that complimented similar text processing skills in ELA.
Measures of Success (Sheldon et al., 2010) contains a preface which required that students learn basic technique (posture, instrument carriage, breathing, and articulation) prior to the introduction of musical notation. This playing technique was applied to four songs, including Hot Cross Buns and Mary Had a Little Lamb, which were written in alphabetical text and were already familiar to many students through aural traditions. This referenced the sound before sight phenomenon in language acquisition common in ELA (Hansen et al., 2014). As students progressed, they strengthened their ability to decode, interpret, and respond to notation, thus transferring the language-learning process to new contexts through derivative subsumption (Ausubel, 2010). In this manner, the theory of cognitive structuralism is practically applied and utilized in the context of a beginning band rehearsal.

Students were taught to tap their foot to help them keep a consistent pulse throughout the piece. This simple process made students physically accountable to timing and rhythm, though students in the study demonstrated varied capabilities to accomplish the task. Research has confirmed the correlation between rhythmic ability and text comprehension, particularly as related to the fluid reading of text silently or out loud (Douglas & Willatts, 1994; Klauda & Guthrie, 2008). Therefore, the explicit instruction of rhythmic accountability through foot-tapping may have had a positive impact on student reading achievement.

Also, this study generated a class environment that not only systematically rewarded risk-taking, but also fostered the application of reading skills in a motivational setting that was collaborative, enriching, and personally fulfilling. Unlike the typical ELA classroom, band participation involves learning in a community whose success depends on the shared growth of its participants. Weren, Kornienko, Hill, and Yee (2017) explored physiological, social, and
evolutionary rationales that explained why students are motivated to participate in group music ensembles:

Some researchers, including Walter Freeman (2000), suggest that participation in social music making by early humans performed an important evolutionary role in the development and survival of societies. McNeill (2000) discusses how the communal experience of music making involves actively creating a sense of unity that facilitates cognitive coordination, shared emotional states, “boundary loss,” and a development of trust in the community of music makers. Others have proposed that group music making modulates oxytocin and other neuromediators and hormones that facilitate social cohesion, bonding, and building trust and strength among members in communities. (p. 320)

Therefore, band participation likely produces a motivational benefit resulting from students’ membership in a socially-based class. Students who do not participate in band may be attaining these benefits in another organized activity or class, and a student participating in band is not guaranteed such motivational benefits. However, students are more likely to experience socially-based motivation as a member of a music ensemble, and in turn, are more likely to participate in the innate literacy practices that are associated with music reading and performing. Motivation and social dependency (Vygotsky, 1934/2012) should not be underrated as an influence on literacy growth.

Several factors may have contributed to the lack of statistical significance despite the parallel processes and theoretical framework previously stated. The primary fault of the study design was low power. Power is the probability that the test will reject a false null hypothesis or the likelihood that the test will distinguish an actual effect from pure luck. Power is affected by the amount of unexplained variance in the data, the magnitude of the effect, and the size of the sample (Field, 2013).
If the current study is replicated in the future, several recommendations may help improve the power. The sample was too small and relatively homogeneous. More participants and greater heterogeneity would allow for much more flexibility and statistical validity, including the possibility of hierarchical modeling or the introduction of a second covariate based on socio-economics or single/married household status, both of which were shown to be significant moderators of academic growth in related studies (Johnson & Memmott, 2006; Kinney, 2008). The sample size might have been improved with different recruitment techniques. Daily reminders through the school-wide announcements, notifications during lunch, and more direct messaging to parents (e.g., through e-mail or postal mail) would have helped improve the number of participants. These options were not easily available at the selected site of the current study.

Additionally, power could be improved by increasing the strength of the treatment (Shadish et al., 2002). This could be accomplished by improving the rigor of the band experience by adding rehearsal time, creating more formalized practice standards, or increasing accountability for assignments (e.g., through report cards, video/audio recordings, or progress reports of some type). A more thorough analysis of the band rehearsals could reveal opportunities to be more efficient with the use of time, including keeping students engaged longer, conducting faster transitions, and utilizing student leadership to accelerate the pace of instruction. A different set of literacy-enriched strategies may have also had an effect. Lastly, a longer intervention would have improved the strength of the treatment. If the study were initiated sooner, students would gain an additional 8 weeks of instruction. However, the logistics of attaining permissions and funding for the study prevented instruction from beginning earlier in the school year.
Some may misinterpret the affirmation of the null hypothesis as proof that the theory is false as applied in this study. This is not the case. Rather, it demonstrated that more study was required to validate the alternative hypothesis. The theoretical framework still maintains that cognitive structuralism was plausible as an explanation of literacy growth among conventional band students, but the design and statistical model of the current study was insufficient to find the evidence necessary to reject the null hypothesis. Therefore, it is hoped that the study may be replicated with the recommended adjustments.

Research Question 2: Interpretation and Discussion

Question 2 asked, “How does literacy-enriched band instruction explain higher textual literacy score means over the absence of band instruction among elementary students when controlling for pretest differences?” Like question 1, question 2 is explained through cognitive structuralism. The only prototypical difference between the literacy-enriched group and the conventional group was the intervention of ten explicit literacy strategies into the lessons and rehearsals of literacy-enriched participants. The ten strategies were chosen among a list of approximately 26 strategies because they correlated with strategies already used by the fourth-grade class, they were easy to implement, and they were deemed to be highly effective in the literature (Daniels and Zemelman, 2014; Hansen et al., 2014; Urquhart and Frazee, 2012; Vacca et al., 2011). A replication of the current study may select different strategies based on the conditions of the testing site.

The strategies may be categorized into strategies appropriate for pre-reading, strategies for use during reading, and strategies for summarizing or reflection. Pre-reading strategies included Frontloading with Images, which helped pique student interest and curiosity in reading
material by pre-alerting them to new information through pictures. Pictures included the method book’s use of composer portraits and a conducting diagram. Pictures also included the introduction of new symbols, such as a fermata, and new pitches, such as concert A, where students were asked to predict what the symbol meant or what the composer was trying to communicate when the portrait was created. Another pre-reading strategy was Vocabulary Predictions, where students attempted to categorize unknown words. Through the categorization process, students learned to associate words with the word group, which helped them to retain the words in their long-term memory. Words included clef, time signature, repeat, fermata, quarter, half, and whole. Additional pre-reading strategies included the Read Aloud and the Think Aloud, where the teacher modeled the thought process of music performance, including identifying the key signature and time signature before playing, investigating unusual patterns of notes or rests, and fingering unusual note transitions prior to the performance. The Read Aloud also allowed students to hear proper inflection, tone quality, breath support, and fluidity. The read aloud was the only strategy that was used in both the conventional and literacy-enriched groups.

Strategies used during reading included the Turn and Talk. Students used this strategy to help describe, evaluate, or summarize new information to their peers. For example, when the instructor introduced key signature, students used the Turn and Talk method to describe where it was located and which notes it impacted. Students also used the Turn and Talk to perform for each other, helping their peers with rhythms, fingerings, and error identification. Other strategies included using Post-it Response Notes, Coding, and Annotating to jot down important ideas during the reading process. For example, if students missed the same fingering more than three times, she was asked to code it (write a symbol), annotate it (write a few words), or write a
message on a post-it note (such as a reminder to practice later or a story to associate the fingering with a particular item). The post-it notes could have been collected later and used to create an archive or memory book at the end of the instructional unit, but time was not available at the end of the study to complete that task.

Strategies used after reading involved reflection or summarization. These were accomplished through the Turn and Talk strategy and through Exit Slips. Ten strategies in total were reserved for the literacy-enriched band group. Although the concert performance may be considered an additional summary strategy, it was not included in this account since it was not unique to the literacy-enriched group. Both groups participated in concert demonstrations of text vocabulary, music notation, and rhythmic fluency.

This intervention of strategies produced a statistically significant probability that the correlation of the literacy-enriched band group with the MAP literature subtest was not due to random or unexplained error. The observed effect size was also calculated to be over one standard deviation from the conventional band group, which was a large observed effect according to related literature (Neuharth, 2000; Wolf, 1986). The statistical significance and effect, combined with the randomized experimental design and corresponding research, provided the conditions necessary to conduct an argument in favor of causation.

Generalized causal inferences in the social sciences are never perfect relationships, but rather “inus” conditions (insufficient but non-redundant part of an unnecessary but sufficient condition, Shadish et al., 2002) that include moderating and mediating variables. A process for analyzing causal relationships was proposed by Shadish et al. (2002) and extrapolated in Chapter 3. Simply stated, the process systematically identifies which operations of the claim are quintessential, which operations are irrelevant, and which operations are detrimental to
generalization. The process investigates both construct validity and external validity over all four operations, which include persons, settings, treatments, and outcomes. Since test results produced extremely large effects and probabilities, the causal claim process was activated to determine whether a generalized causal inference could be inferred from the data.

**Literacy Tests: Surface Similarities**

The prototypical operations (persons, settings, treatments, and outcomes) of the study were clearly defined in the interest of making a generalized causal inference. These constructs constituted the essential elements that were germane for reproducing the causal effect in practical field applications. The prototypical characteristics will be defined to establish construct validity, then acceptable variants will be identified to help establish external validity. Many of these characteristics were defined in detail in Chapter 3 and are summarized here. In the absence of related experimental studies, non-experimental studies will be used to contextualize the findings and strengthen external validity.

The persons consisted of non-special education elementary students with no prior band experience (beginners). Since band may begin in grade five or higher in some districts, an acceptable variant would be to alter the grade level to include beginning band students through eighth-grade (see Baker, 2011; Huber, 2009; Johnson & Memmott, 2006; Kurt, 2010). The setting was a public school in Illinois. Variants may include public schools in any region of the United States, since related non-experimental studies have been performed in Illinois (Wachtel, 2006), South Dakota (Neuharth, 2000), Florida (Barlar, 2010), Ohio (Fitzpatrick, 2006), New York (Huber, 2009), and Iowa (Kurt, 2010). In other words, since related research has found
similar results among the same identified characteristics of persons and settings, the causal claim of the current study is strengthened to include only those characteristics.

The treatment consisted of administering the conventional and literacy-enriched teaching strategies exactly as indicated in Appendix F. The treatment occurred during one lesson and one full band rehearsal each week for 14 total sessions. The lesson lasted 30 minutes and the rehearsal lasted 45 minutes. Furthermore, the instructor was an experienced teacher with at least five years of experience teaching beginner band students. The literacy-enriched group differed from the conventional band group with the addition of a finite number of explicit literacy strategies (10) to the conventional strategies. At least one strategy was utilized at each full band rehearsal, and then explored and applied during the small group lesson. Variations may include minor lengthening or shortening the instructional time, but lessons less than 18 minutes were deemed not effective for revealing change effects (Kemmerer, 2003). Variations on similar student-centered instructional techniques were also advocated and detailed by Henninger, Flowers, and Councill (2006), Miksza, 2013, Heuser (2011), and Kratus (2007), but little research exists considering their comparative efficacy (see Bazan, 2011).

Lastly, the outcome was investigated for surface similarities with related studies. It consisted of literacy growth as measured by the overall MAP reading test and its subtests as defined by the NWEA (2017a). While the reading test measured overall comprehension, the literature subtest only measured students’ ability to evaluate and identify the structure of fiction texts. The MAP was administered the week following the last lesson. Variants to the outcome include administering the test in different times, locations, or seasons, assuming the lag time after treatment is consistent with the current study. Variations represent a wide heterogeneity regarding the delivery and timing of the posttest (e.g., Gromko, 2005; Neuharth, 2000; Barlar,
2010), which indicate that flexibility regarding the delivery and timing of the test is an acceptable variation from the outcome of the current study. Thus, the outcome contained shared surface similarities with related studies, thereby affirming the claim for a causal inference.

**Literacy Tests: Rule Out Irrelevancies**

Several characteristics were deemed irrelevant for the purpose of replicating the study. Though an infinite number of variables influence causal relationships, some variables are seemingly independent of the causal web. Further research may find different irrelevancies or find that some irrelevancies are actually moderating or mediating conditions, which is an expected result of continued inquiry on the topic (Shadish et al., 2002). The following operations are an acceptable best guess for irrelevancy, based on available evidence from this study and related studies.

The persons involved in the study demographically varied on an unlimited number of characteristics. Four characteristics were tested and shown statistically to be irrelevant—gender, socio-economic status, ethnicity, and classroom teacher. Data were also collected on age, prior music experience, mobility rate, single/dual parent household status, and potential transportation issues; these data were not sufficient for a full statistical analysis due to the homogeneity of component groups. A few divergent cases within the variables of age, prior music experience, mobility, or transportation needs did not constitute statistical outliers within the sample pretest or posttest. For example, two students took private music lessons outside of school. However, both studied less than a year and neither scored far from the group mean pretest or posttest. It can be concluded that these characteristics were not relevant in this study. No other related studies reported on the relevancy of the participants’ internal (within-group) homogeneity or
heterogeneity. Still, studies with a more heterogeneous sample have shown ethnicity, parent involvement, and SES to be relevant among heterogeneous populations (Johnson & Memmott, 2006; Kinney, 2008). Therefore, they will be discussed as discriminations in the following section.

Regarding the setting, the exact location of the urban school was determined to be irrelevant. Many studies have investigated similar urban school settings and found significant differences among band and non-band participants, such as the urban school setting of Kinney (2008) or Costa-Giomi (1999). Furthermore, lessons and full band were taught in the cafeteria, as the school had no band room. Instructional locations were deemed not relevant or not addressed in related non-experimental literature.

The treatment contained several irrelevancies in the implementation of the study. These included the time of day that lessons were delivered, the method book used, instruments selected, and slight variations of instructional pacing and nuance as necessary for individualized instruction. The time of day was irrelevant because lessons were taught both before and after school in this study, yet neither group scored significantly differently from their conventional or literacy-enriched group mean. The method book was not a contributing factor in related non-experimental studies, as evidenced by the overall lack of identification or even mention of method book usage in the literature. Notably, several articles compared method books to each other to examine content (see Brittin & Sheldon, 2004; Byo, 1988), but not as a factor relating to academic development. Furthermore, teachers are typically allowed the flexibility to slightly vary their instructional pacing depending on the needs of the individual student. To be clear, the week-to-week pacing and strategies were consistent, but the internal pacing, points of emphasis, and differentiated treatment of individual students was flexible, as it would be in any similar
field experience. The instruments consisted of flute, clarinet, trumpet, trombone, and drums, which were the same instruments used in the curricular program at the fifth-grade level at the participating school. The isolation and examination of instrument selection as a confounding factor showed no significant differences on posttest scores. Lastly, this study utilized the NWEA MAP literature subtest as an outcome, but related research suggests the test type was irrelevant to the examination of literacy growth. Other norm-referenced tests have produced similar results, such as the Comprehensive Tests of Basic Skills (Neuharth, 2000), the Iowa Test of Basic Skills (Kurt, 2010; Wachtel, 2006), and the Illinois Standards Achievement Test (Wachtel, 2006). Several criterion-referenced tests have also shown similar results, such as the Pennsylvania System of School Assessment (Thornton, 2013), the Louisiana Educational Assessment Program (Baker, 2011), the Ohio Proficiency Test (Fitzpatrick, 2006), and the New York State Testing Program for English Language Arts (Huber, 2009).

**Literacy Tests: Discriminations**

Discriminations identify the characteristics of the operations that are not generalizable. Student’s ethnicity and SES status were internally homogeneous, so they did not impact posttest scores in this study and were shown to be irrelevant locally. But externally, the sample did not feature typical demographic proportions of ethnicity or SES among 4th graders in the United States (NCES, 2017). About 40% of the participant families spoke Spanish as their primary language, which also reduced generalizability. Additionally, the sample represented only about one-third of the total fourth-grade population. By asking students to return study permission forms, the study design required a form of volunteerism that may have impacted the outcome (Shadish et al., 2002), despite efforts to recruit a large heterogeneous sample. One clear
advantage of the study was that band enrollment was typically around 15 students in fifth grade, and this study recruited 43 students, of which 32 were included in the analysis. Therefore, the current study included data from many students who otherwise would not have participated in the curricular band program, which improved the external validity of the sample.

Regarding the setting, the urban location of the school may reduce generalizability when compared to suburban or rural locations, where parental involvement, cultural norms, and school support may be different. In other words, generalizability may be limited to similar urban settings. The support or non-support of arts in private, rural, or suburban schools often depends on factors that are yet unknown (Abril & Gault, 2008; Ravitch, 2014). Plus, the environmental impact of a dual language classroom on an otherwise mono-language grade level has shown mixed academic impacts in the literature (see Lindholm-Leary & Block, 2009; Valdez, 1997). Other threats to generalizability include the treatment sequence design, which was created in consultation with fourth-grade teachers to align with the timeline and readiness of the fourth-grade class at the selected school. Other schools may have different timelines and may emphasize different literacy strategies. Lastly, regarding outcomes, the MAP test did not test many aspects of the comprehensive literacy model presented in Chapter 2. The test was limited in its capacity as a comprehensive norm-referenced exam, and only the literature subtest produced significant results. The remaining reading tests may require more time or may test separate skills. Other testing instruments would be necessary to investigate literacy components that were important to literacy achievement but not measured on the MAP literature subtest. Generalizations noted among the irrelevancies may be limited to the components that are similar among the literature subtest and related criterion- and norm-referenced tests (such as the identification of theme, structure, plot, and other literary elements).
Interpolations and extrapolations are necessary to help define the scope of the causal claim. Although not all possible values were observed on the continuously measured MAP posttest RIT scale, the results may be interpolated proportionally similarly to the results that were observed and recorded in this study (NWEA, 2017a). For example, a RIT pretest score of 200 can be assumed to have a proportionately similar posttest score between 198 or 202, even though a score of 200 was not observed within this data. Due to the reliability of the RIT rating system (Northwest Evaluation Association, 2017b), this interpolation is an acceptable assumption of the study.

Extrapolation is easier because many studies have examined correlations that extend into the age and time range of the current study. For example, it may be assumed that a longer treatment will produce greater effects, due to multi-grade level studies by Huber (2009), Kinney (2008), and Thornton (2013). More instructional time per week and more weeks per program would have similar effects (Gromko, 2005; Kemmerer, 2003). Therefore, data may be interpolated and extrapolated proportionately, as indicated by related literature providing similar results.

The foundational theory that binds together the surface similarities, dismisses the irrelevancies, accepts the discriminations, and accommodates the interpolations and extrapolations is the theory of cognitive structuralism (Ausubel, 2010; Bruner, 1977, 1987, 1997). Derivative subsumption is an important component within cognitive structuralism, and it
explains the interaction of previously learned concepts into new cognitive structures. For example, if students practice predicting vocabulary in several contexts such as music and ELA, they will internalize and apply the concept more intuitively when it is presented on an exam. This phenomenon explains the high effect size and the non-zero-spanning confidence interval indicated by the literature subtest score.

Whereas the 10 literacy strategies produced improved growth in both experimental groups, the greatest growth occurred in the literacy-enriched group on the literature subtest. This is not surprising since most of the reading strategies presented to the literacy-enriched group focused on systematically noting, recalling, and interpreting main ideas in the text or in the music. For example, students annotated and coded when they stumbled over a difficult note, they used visual cues (pictures) to help interpret new notes and provide context to a work’s historical meaning, they practiced thinking and discussing performance strategies with their peers using the turn-and-talk strategies, and they used post-it notes to identify questions or important material. These strategies were learned through repeated practice, and the addition of the musical context provided the structure necessary to strengthen their skills and produce much greater results than would otherwise be observed.

Alternative accounts could not explain these effects with the same certainty. The randomization of all consenting participants helped distribute confounds among all three groups, whether they were known or unknown. Tests rejected the hypothesis that groups were differentiated by gender, SES, ethnicity, and classroom teacher. Instrument, age, parent involvement, private music lesson involvement, mobility rate, and transportation requirements were also determined to be irrelevant. The setting and testing protocols were identical between the two groups, and non-study related literacy instruction was divided equally among all three
groups. The MAP test has been independently tested for reliability and validity (NWEA, 2009). Six students who did not adhere to their randomized assignment were excluded from the analysis. Although unknown confounds may be responsible for producing the significant findings, none were found in related literature. Though it is possible that a hidden confound is responsible for influencing the statistically significant results, it is highly unlikely.

One alternative explanation suggests that the Hawthorn Effect was responsible for the improvements. This source of bias occurs when participants in experimental groups react differently because they were aware of their participation in a study (Cohen, Manion, & Morrison, 2011). Resources were not available to create a blind design or use placebos. Though students were aware they were involved in a study, the experimental groups were treated equally and without bias. Neither the instructor or the researcher knew whether one group was better than another on any factor, so neither group received preferential treatment. However, the Hawthorne Effect remains a valid critique and is accepted as a weakness of this study.

Lastly, statistical conclusion error was possible, especially considering the high number of tests performed. To help mitigate statistical error issues, multiple indicators of probability and effect size were calculated. The alpha level and beta level were set a priori, and adjustments were anticipated and performed as a result of the low sample size. A Bonferroni adjustment helped alleviate a concern of data fishing. Efforts were made to build power by reducing the random error, such as utilizing an ANCOVA instead of a repeated-measures ANOVA, increasing the alpha level, and investigating one-tailed tests instead of two. These were acceptable accommodations given the sample size, hypothesis, and lack of corroborating causational literature in the social sciences (Cohen et al., 2011; Shadish et al., 2002). Furthermore, the
reporting of confidence intervals, standard error, and effect size helped readers contextualize the results and draw their own conclusions if they choose not to accept the chosen alpha level.

The final argument advocating for a causal explanation is the comparison of effect sizes among related studies. Although this study did not attempt to perform a meta-analysis, the effect sizes observed in related studies helped provide a context of the magnitude of the differences observed in the current study. Wolf (1986) accepted Cohen’s $d = .25$ as an educationally practical value, and $d = .50$ as a clinically practical value, but he added, “it is better to obtain these standards for comparison from the professional literature than to use these somewhat arbitrary guidelines” (p. 27). With this in mind, quasi-experimental studies relating music instruction with overall academic achievement (including the band and ELA studies cited in Chapter 2) had a mean $d$ of .265. Causal comparative and correlational studies had a mean $d$ of .272. The results of the related literature concur with Standley’s (2008) meta-analysis mean $d$ of .250 for evaluating music’s effect on reading scores. Therefore, an approximate $d$ of .250 was to be expected. However, the actual value was far greater. The effect of the literacy-enriched group on the literature subtest in the current study was more than four times Standley’s (2008) value at $d = 1.09$. The observed $d$ value represents both an extremely high practical value and a high magnitude of change.

Like the first research question, however, the remaining three reading tests (overall, vocabulary, and informational texts) provided no statistical evidence to show how literacy-enriched band instruction caused higher textual reading scores. There were many possibilities to explain why the literature subtest was significant while the remaining tests were not. Since the literacy strategies were not differentiated according to the test categories, it was difficult to attribute the difference of achievement on the types of literacy strategies used. However, the
generalized transference of text literacy skills to music notation involved higher-order processing skills that may have benefitted the literature subtest objectives more than the other tests. The abstract association of symbols with sounds may have helped band students better visualize, comprehend, and interpret works of fiction as required by the subtest. Certainly, further study may help identify which salient qualities of the literacy-enriched group best caused literacy growth, including the use of mixed-methods and qualitative designs. Nevertheless, the theoretical explanation and research-based strategies of the current study may still provide utility to researchers and band teachers seeking to expand their pedagogical skills.

Research Question 3: Interpretation and Discussion

Research question 3 asked, “Does literacy-enriched band instruction yield a different effect from conventional band instruction on textual literacy scores among elementary students?” Research question 3 was not an explanatory question as were questions 1 and 2, but rather an exploratory question seeking to find whether one type of band instruction was superior to the other regarding text literacy scores. The ANCOVA revealed that the four reading test scores were not statistically significantly different between instruction types; therefore, the type of instruction was found to be irrelevant in its impact on textual literacy achievement. This finding was notable since the literature subtest revealed a statistically significant increase from the control group to the literacy-enriched band group. However, the difference between the conventional band group and the literacy-enriched band group was much smaller in effect and power, thereby failing to reach the critical value necessary for rejecting the null hypothesis.

The two groups shared many similarities that may be ruled out as moderating variables in the correlation of the two groups. These include a completely randomized group assignment
which distributed potential confounds, known and unknown, probabilistically across both
groups. The setting and teacher were the same for both groups. The treatment featured the same
method book for both groups, and students progressed through the book at roughly the same rate.
Instructional time was equal, as was non-study related literacy instruction across the fourth-grade
classrooms. Students who switched groups were not factored into the tests, eliminating a
possible source of bias. Lastly, the outcome instrument was the same for both groups, as both
groups were administered the MAP among identical testing conditions.

Differences between the two groups included the day of the week of instruction and the
type of instruction received, as differentiated in Appendix F. The instruction type comprised the
addition of 10 explicit literacy strategies to the conventional strategies used by most ELA
teachers, such as vocabulary predictions, think alouds, turn and talks, summarizations, and
annotations. Other variables may have influenced the outcome, but were hidden, latent, or
unknown. The amount of random error suggests that such variables existed.

Several explanations were explored to investigate why the conventional band group and
the literacy-enriched band group failed to differ on MAP reading scores, despite the strength of
the cognitive structuralist theory and the oppositional findings of other quantitative literature.
For example, the Hawthorne Effect could have influenced the outcome, similar to research
questions 1 and 2. Students who were aware of their experimental conditions sometimes
perform differently (Cook & Sinha, 2006); it is possible that literacy-enriched students were less
motivated or conventional students were more motivated as a result of their group assignment.
This criticism is noted, and the only solution was to utilize a placebo or blind study, which was
beyond the time and resources available for the current study.
Another explanation suggests that threats to construct validity may have influenced the outcome. Specific threats included instructor expectancies, when the instructor intentionally or unintentionally conveys higher expectations towards one group; compensatory equalization, when the instructor intentionally or unintentionally provides extra treatment to the non-treatment or less-treatment group; and treatment diffusion, when participants receive interventions intended for the other group (Shadish et al., 2002). One literacy strategy was identified as threatening treatment diffusion—musical read alouds. The musical read aloud occurred when the instructor sang or played for students to illustrate fluency and inflection. However, the strategy was applied equally to both groups so that any effects resulting from the strategy were equally distributed and thus eliminated as a confound. A textual read aloud was performed once for the literacy-enriched group (lesson 14), but it was not performed for the conventional group, and thus contributed as one of the ten differentiating literacy-enriched strategies.

The lack of similar experimental studies necessitated the use of non-experimental studies to contextualize the findings for question 3 in the literature. For example, Bazan’s (2011) two-part mixed methods quantitative design found that bands that behaved well were more likely to engage in student-centered activities. Although the current study did not investigate behavior as a mediating variable, it shared the finding that student-centered, progressive behaviors tend to produce growth in both musical and non-musical areas. Similarly, Heuser’s (2011) qualitative approach found that conventional instruction, rooted in competition and external motivation, was not as effective or meaningful than student-centered, progressive instruction.

In summary, the posttest score differences were controlled by pretest scores and the confounds were minimized so the resulting means differences could reasonably be correlated with the treatment of literacy-enriched instruction over conventional instruction in this study.
However, none were statistically significant. Therefore, these findings fail to reject the null hypothesis that means were statistically different between the conventional group and the literacy-enriched group. The lack of statistical significance justifies the need for further study due to the strength of the theoretical construct, the number of related studies providing evidence to the contrary, and the low power of the current methodological design.

Research Question 4: Interpretation and Discussion

Research question 4 asked, “Does literacy-enriched band instruction yield a different effect from conventional band instruction on musical literacy scores among elementary students?” Literacy-enriched instruction requires time, and most conventional approaches do not easily accommodate the flexible schedule required by blending progressive instruction with conventional instruction. Considering the theory of cognitive structuralism, literacy-enriched instruction should lead to increased musical literacy, and as a corollary, musical performance. Since many of the physiological process of reading and interpreting musical symbols are similar (Hansen et al., 2014), the systematic and explicit use of literacy strategies in music reading should improve the accuracy and fluency of music reading the same way such strategies improve the accuracy and fluency of text reading. However, many directors question whether the investment in student-centered, literacy-enriched instructional time produces significant gains in musical ability (Bazan, 2011).

Students from both groups completed the Watkins Farnum Performance Scale to assess their relative musical abilities after 14 lessons and 14 rehearsals. The conventional group ($M = 9.91$) scored nearly identically to the literacy-enriched group ($M = 9.39$) with an effect of $d = .083$, which was not educationally, clinically, or statistically significant ($p = .850$). Therefore,
the test failed to reject the null hypothesis that the means were different between the conventional group and literacy-enriched group. The findings indicate that time invested in explicit literacy instruction did not produce significant gains compared to the conventionally-taught band group. Perhaps more importantly, the time invested in literacy strategies did not detract from the musical experiences of students. This finding is consistent with Bazan (2011), who found that student-centered strategies (among which literacy-enriched strategies are included) were possible to implement without detracting from individual playing performance goals.

If the null hypothesis was true in the general population, then one must ask to what extent can literacy-enriched strategies benefit band students in more progressive, student-centered ways, without diminishing performance-based results? Most of the content-area literacy research is focused on transforming teacher-centered, autocratic instruction to more student-centered, democratic instruction (Daniels & Zemelman, 2014; Urquhart & Frazee, 2012; Vacca et al., 2011). Student-centered instruction is advantageous for long-term understanding because the material is more relevant, adaptive, and reflective of student responsibility and empowerment than conventional models, which are overwhelmingly performance-based (Miksza, 2013). It would be therefore advantageous to maximize student-centered instruction as long as performance goals are not compromised, and the current study shows that this is possible.

Notably, the current study did not neglect a performance-based product, as both band groups performed a concert for the public. However, the concert in this context was considered an informance (see Nowmos, 2010) or a demonstration concert, where the audience is introduced to the sequence of learning experiences for beginning band students rather than a formal presentation of literature. Students demonstrated how they sounded at various stages of
development, including beginning sounds—brass mouthpiece buzzing, flute headjoint tones, and clarinet mouthpiece squeals. This type of presentation is common in the general music realm (Kerchner, 2010; Nowmos, 2010), and band teachers can benefit from following the model. At a more advanced level, the concert type would impact the criteria of the literature chosen and ultimately lead to richer, more culturally and historically diverse pieces that address more of the left side (background knowledge) of the literacy model presented in Chapter 2. The concert then becomes a venue for educating the audience about the context, structure, purpose, and important features of the pieces performed. The literal recreation of the composer’s notation is only one of many goals in a public performance. This type of presentation is congruent with all facets of the holistic literacy model presented in Chapter 2.

If the null hypothesis is false in the general population, then one must ask what unanticipated explanations may have contributed to the non-significant findings of the current study? The lack of a means difference between the two groups could have been caused by the short time span within which the intervention was administered. Fourteen lessons and 14 rehearsals represent a course of about half of a typical school year. True benefits of literacy-enriched instruction may not occur until after one or several years of band instruction. Huber (2009) maintains that such effects may take up to two years before becoming detectable. Related research has shown student-centered instruction to be advantageous for long-term musical growth (Heuser, 2011; Kratus, 2007). Therefore, a longer treatment may provide different results.

Another alternative explanation suggests that the specific strategies used were not differentiated enough from conventional instruction to observe a large effect. Table 2.1 presents 26 strategies that are mentioned in content-literacy texts, so perhaps another combination of
strategies would produce different results. This explanation is countered by examining the results from the text literacy portion of this study (within research question 2), which did find that the literacy treatment was strong enough to produce at least one significant result among the literacy scores. Therefore, the treatment had the potential for obtaining the necessary power, but the effect difference simply was not present in the data.

Lastly, an alternative explanation suggests that teacher expectancies may have reduced the effect on the literacy-enriched group, while compensatory equalization may have boosted the achievement of the conventional group. These critiques are valid, but highly unlikely based on the results of the text literacy portion of this study, which did show significance. If teacher expectancies and compensatory equalization were present, the text literacy scores would also have been equal, which they were not.

Evaluating the Theory: Cognitive Structuralism

The theory of cognitive structuralism plausibly explains why so many quasi-experimental, causal comparative, correlational, and meta-analysis studies have found that music reading and text reading are related. However, studies involving beginning band students have failed to show that connection in experimental settings. The current study is the first to make the connection, but it is also the first experimental study in the literature. More research is needed to confirm the applicability and strength of the theory in experimental settings.

Nonetheless, several components incorporated into the design and implementation of this study reveal a strong foundation of cognitive structuralist principles. Conventional and literacy-enriched band students started their instruction by associating the physical movement and the manipulation of the instrument to their aural memory of early childhood folk-tunes, thus
demonstrating derivative subsumption. As students learned new notes, they were comparing and contrasting fingerings, lip pressure, body carriage, pitch, and visualizations. This process asked students to relate old knowledge to new knowledge, thereby demonstrating correlative subsumption by affirming and strengthening the old knowledge.

These components alone did not cause significant growth in text reading scores. However, when combined with literacy-enriched techniques, significant growth was demonstrated, but only on the literature reading subtest. Care must be taken to acknowledge that the processes of delivering the treatment and analyzing the outcomes were complex and subject to variance simply due to human nature. However, the magnitude of the growth was very large and the probability of error was very low, illustrating that the likelihood that the theory of cognitive structuralism being false in the population was relatively small.

Literacy-enriched strategies that supported the cognitive structuralist theory were not specific to each subtest, meaning that the 10 literacy-enriched strategies chosen were not strategically correlated with a test or subtest. However, the combination of strategies applied in the current study did produce unique growth on the literature subtest. Therefore, one may reasonably assume that a different combination of literacy-enriched strategies may produce growth on a different subtest or set of subtests.

Table 2.1 presented 26 strategies that were mentioned in multiple content-area literacy textbooks. All of them support cognitive structuralist principles in two ways: they draw on students’ prior knowledge, and they strengthen students’ ability to process information in multiple contexts, particularly when the same strategies are applied in both ELA and band. For example, the learning log (Vacca, Vacca, & Mraz, 2011) asks students to document and reflect on their learning in the same way that a practice log is utilized in band. The word wall (Daniels
& Zemelman, 2014; Hansen, Bernstorf, & Stuber, 2014) allows students to learn and recall vocabulary in both ELA and band by reinforcing similar concepts of association. The Socratic seminar (Urquhart & Frazee, 2012) allows students to take ownership of their learning in both band and ELA contexts through metacognition, thoughtful peer interactions, and thorough understandings of the content. Although this represents only a few of the strategies, they all demonstrate how cognitive structuralism supports and explains the reciprocal relationship between band and ELA.

Therefore, within the context of this experimental study, the theory of cognitive structuralism is affirmed as it relates to the relationship between literacy-enriched band instruction and the MAP literature subtest. However, the data is insufficient to draw causal relationships among the remaining groups, so the theory is not yet affirmed for band students in other measures of literacy. Since there are so many literacy-enriched strategies from which to choose, there remains a possibility that another combination of strategies may produce significant relationships without any further adjustments to the design or statistical model. Nonetheless, incorporating slight adjustments to the design setting, enhancing the participant recruitment techniques, and employing more robust statistical tests backed by larger samples produces a truly intriguing opportunity to affirm the theory in future studies, considering that the preliminary evidence is so strong.

Limitations

Several factors limited the findings of this study, particularly among construct validity and external validity claims. For example, without any experimental studies from which to compare, the comparison of effect sizes and study operations were tenuous. It is hoped that
future research may provide a larger pool of data to help contextualize the significance or non-
significance of these findings.

Additionally, complexity theory marginalizes all quantitative findings due to the
uncertainty and unpredictability of social interactions (Cook & Sinha, 2006). Complexity theory
is especially applicable when discriminating teaching paradigms. Although teachers may favor
one paradigm over another, they rarely exclude a paradigm entirely (Miksza, 2013). Rather,
teachers borrow from one or another paradigm as needed. Furthermore, the magnitude and
quality of the instructional strategy matter, and neither were controlled in this study due to the
time and resources available. This concern was attenuated through the use of lesson logs and
thorough planning. In such circumstances, causal inferences are tenuous, but still possible.
Causal inferences require statistical logic, sound explanatory theory, and the explicit definition of
constructs (Shadish et al., 2002). Causal inferences are also strengthened by developing a body
of confirmatory studies, which are investigated through meta-analysis. As of yet, neither
confirmatory studies nor meta-analyses exist to strengthen the causal inference. For this reason,
all causal inferences must be applied with caution.

Furthermore, Shadish et al. (2002) recommend randomizing all operations of a study to
truly eliminate all spurious factors. However, resources almost never exist to create multiple
operations (heterogeneous populations, settings, treatments, and outcomes) from which
researchers may randomize. The current study followed a more typical experimental design
where participants were randomly assigned to groups. But despite efforts to create such an all-
inclusive and unbiased sample, participants still comprised a subset of volunteers because
consent to participate was required. Volunteers may have behaved differently than the targeted
population. This criticism was mitigated by randomizing those who volunteered to participate.
Therefore, students who volunteered were compared against other volunteers, obtaining the intended effect of dispersing confounds equally at the cost of generalizability.

Another limitation concerns the construct validity of the outcomes. The outcomes observed did not represent a holistic measure of textual literacy or music literacy. The tests were internally valid according to the literature (NWEA, 2009; Watkins & Farnum, 1954), but were only tangentially valid in their ability to measure the overall effects of the treatment in the broader sense of holistic literacy and cognitive structuralism proposed by Bruner (1977, 1987) and Ausubel (2010). The MAP tested several distinct aspects of textual reading, which constituted a small subset of the overall textual literacy model. The WFPS was a sight-reading test, which represented a small subset of the overall musical literacy model. Neither test claimed to be a comprehensive literacy test. The criticism was acknowledged, but a more elaborate literacy assessment was beyond the scope of this study. It is possible that students grew in their literacy skills and musical skills in ways that were not measured by the selected tests. However, since all students were assessed fairly and equally, the tests provided useful evidence for comparing literacy growth within the limited scope indicated by each test. In that regard, the tests served their intended purpose.

A final limitation consists of a philosophical debate concerning the nature of aesthetic education. Leonhard and House (1959), Reimer (1970), and Eisner (2001) were proponents of the aesthetic justification of music education, which validates the study of music on its own artistic merit. They denied the rationalization of music study through the lens of another academic subject. Their criticism was appropriate, as the study of music can never replace the study of literacy or any other subject. However, their argument, though simplified here, largely assumes a one-way relationship. The current study, on the other hand, assumes that the study of
text literacy can bilaterally benefit the study of music. This concept is especially applicable in beginning band environments, where the content is overwhelmingly focused on reading (Cavitt, 2003; Colson, 2012; Henninger, Flowers, & Councill, 2006; Heuser, 2011; Kratus, 2007; Sheldon, Balmages, Loest, & Sheldon, 2010).

The reciprocal relationship between ELA and band, as justified by the cognitive structuralists, suggests that the aesthetic benefits of music instruction can coexist with the pragmatic benefits of text instruction. For this to happen, the prototypical beginning band model must change. Text literacy strategies may be used to help students lead to more accurate note-reading at the beginning stages, which facilitates greater musical development in the intermediate and advanced stages. At this point, aesthetic experiences are more likely to occur.

The prototypical band model is but one path among many paths that lead to an aesthetic enjoyment of music-making (see Miksza, 2013). Ideally, the elimination of arbitrary divisions between ELA and music enable the justification debate to be refocused. The debate no longer becomes whether music is taught for aesthetic or utilitarian purposes, but how might utilitarian justifications best lead to aesthetic understandings? In this spirit, the current study assumes that in a typical, contemporary elementary school, an aesthetic education in band begins with a focus on literacy. This education should not merely focus on reading but on all elements within the comprehensive literacy model presented in Chapter 2, including composing, improvising, performing, and learning the contextual knowledge necessary to create meaning from the symbols that represent musical notation. This scenario recognizes the complexities inherent in supporting beginning band students in a formalized school setting while acknowledging that typical band instruction consists of one of many paths to attain the goal of experiencing aesthetic beauty.
Implications for Practice

The results of this study produced several implications for practice in the field of education. For example, the separation of ELA and music strategies represent artificial divisions with the same learning process. Band students improved their literacy scores using both conventional band techniques and literacy-enriched band techniques, although statistically significant gains were observed within the literacy-enriched group. This evidence shows that band instruction can cause higher literacy scores in certain circumstances. Legislation or curriculum development that directly or indirectly encourages the separation of subjects (e.g., decontextualized testing) to compare schools or districts creates artificial barriers between mutually beneficial relationships. Those barriers prevent cognitive structures from building under disparate contexts like music and ELA.

Furthermore, teachers must collaborate to eliminate barriers within their own professional communities. Even though the current study is delimited to the study of band, many of the non-performance related elements in the comprehensive literacy model are addressed in general music. In-depth studies in Kodály, Suzuki, Jaques-Dalcroze, Orff and CMP techniques reveal experiences that involve instrumental performances that require no reading, yet demand specific knowledge of music literacy as previously defined. Many band directors have little knowledge of general music methodologies or philosophies (Liperote, 2006), and therefore miss an important opportunity to provide a balanced, holistic music education experience to their students by working collaboratively with general music teachers. Band teachers must utilize the resources available to them, including their general music counterparts, to broaden their pedagogical repertoire (e.g., Mason, 2012).
Correspondingly, band method books must recognize multiple strategies toward building music literacy by incorporating more explicit strategies borrowed from ELA and content-area literacy experts. Most method books miss the opportunity to connect with other disciplines, despite encouragement from the Common Core Standards (CCSS, n.d.) and the National Core Arts Standards (n.d.). The next generation of band method books will certainly reflect a more holistic view of literacy.

Evidence also shows that the practice of pulling students out of band could cause more harm than good regarding literacy development in some circumstances. For non-special education students, band provides the disparate context necessary for building and strengthening literacy-related concepts through correlative and derivative subsumption (Ausubel, 2010). Since the physiological process of reading and interpreting symbols are similar among music and text, removing or denying an opportunity for non-special education students to participate in band may be counterproductive since the opportunity to apply skills to disparate contexts is negated. An exception may include pulling a student from band to participate in a specialized or individualized literacy treatment or remediation program, as this study did not encompass that scenario.

Band teachers must coordinate with ELA teachers to capitalize on parallel strategies, themes, terminology, and sequencing. Similarly, ELA teachers and reading specialists should look to alternative contexts for applying literacy strategies. Teachers of all disciplines have worked in a culture of isolation for many years (see Schmoker, 2006). This culture is shifting under the advent and spread of Professional Learning Communities (DuFour, 2016) and the advocacy for professional collaboration under Domain 4 of the Danielson (2013) framework. As teachers collaborate more, they have additional opportunities to coordinate their instruction.
between subject areas, and students will be more likely to make cross-curricular connections for the mutual growth of all subject areas.

Band teachers may also benefit from adopting more student-centered, literacy-enriched strategies into their regular practice. These strategies may be applied without detracting from the performance goals of the program, which was also Bazan’s (2011) conclusion. By utilizing these strategies, band directors are addressing more of the holistic conceptual literacy model components presented in Chapter 2 which are traditionally ignored or marginalized in American schools. Music should be addressed in terms of its molar parts, emphasizing not just reading and performing, but also writing/composing, listening, and background contextual knowledge. Adopting more literacy-enriched strategies and more student-centered strategies will make learning more relevant and meaningful to students (Kratus, 2007). Advocates for the Comprehensive Music through Performance model have been promoting such a change, and their efforts have resulted in a wealth of information, resources, and strategies to help band directors implement innovative student-centered lessons that reinforce all areas of the literacy model, not just reading (ILCMP, 2014). Use of these strategies has the added effect of supporting and demonstrating several indicators of quality teachers within the Danielson (2013) framework for professional practice.

Lastly, the findings of the current study underscore the need for additional in-service and pre-service development on literacy techniques. College administrators should reconsider the role of explicit literacy instruction in music education teacher preparation programs, since text literacy and music literacy share many of the same pedagogical strategies. Classroom teacher and reading specialist candidates would also benefit from required instruction in content area literacy strategies, which encompasses the application of literacy strategies to all non-ELA
courses such as music, art, science, and technology. Many teacher preparation programs recognize the value of world languages as an important curricular component of a liberal education, with both cultural and linguistic benefits relating to the epistemology and structure of the English language. The same argument may be made for incorporating a performance music component into the teacher-preparation curriculum. A music performance class which includes a literacy component would be a crucial step toward helping general education teachers to learn and apply literacy strategies in a unique and effective context. This requirement would help reduce the individualism that artificially and unnecessarily isolates many early-career band directors, reading specialists, and classroom teachers (Boreen & Niday, 2000). Professional development for experienced teachers would also benefit students, as the collaboration between all staff members is a prerequisite to providing students with a more efficient and more effective educational experience.

Recommendations for Research

The findings fomented several recommendations for future research. First, the duration of the current study revealed a need for a long-term study to investigate longitudinal effects. The amount of time to typically see results from reading interventions ranges widely from just a few sessions (Meyer & Felton, 1999) to two years or more (Kamps et al., 2008). Since causal comparative studies have the ability measure group differences over a span of years, they may be used to help gauge how much time a significant effect may require to be observed. Musical studies in this category require about two years (Huber, 2009) to three years (Piro & Ortiz, 2009) before effects are observable.
A longitudinal study designed similarly to the current study would be possible without incurring significantly higher costs. The researcher could recruit a school district (or several districts) where band begins in grade six or later and then implement an intervention beginning in grade four. This avoids conflicting the experimental treatment with existing curricular band programs. The instruments and books need only be purchased once, while the instructional stipend may be negotiated, funded through a grant, or traded for reciprocal instruction, lessons, or other non-curricular duties. Furthermore, additional test data provided over time will allow an analysis of whether the observed effects diminish or increase with band experience, creating a curvilinear prediction model and creating additional points of interpolation, which builds credibility for the causal inference.

Second, more studies are needed to help identify, define, and examine possible confounds as part of a larger statistical model. Although the current study collected information on gender, age, SES, parent support, instrument, transportation needs, and mobility rates, the data were not heterogeneous enough to examine without bias. With larger samples, however, more statistical options are available, and their findings would be more reliable. Examples may include causal modeling, structural equation modeling, and covariance structure analysis, which benefit from sample sizes above 300 participants (Shadish et al., 2002). A larger sample size would also allow the use of hierarchical or multilevel linear modeling, which would permit sites to be randomly chosen from a larger selection of possible public schools, increasing external validity (Tabachnick & Fidell, 2007). For a truly randomized study, all operations in this study (participants, settings, treatments, and outcomes) could be randomized, entered into a more complex model, and further investigated provided there is enough power. Such a model is exceedingly rare and highly unlikely due to the cost and time involved (Shadish et al., 2002). If
funding and time were available, this model would not be impossible, and would be an incredibly useful addition to the music education literature.

Third, the Hawthorne effect occurs when participants react differently when they are aware of their membership in an experimental or non-experimental group (Cook & Sinha, 2006). A future study could help attenuate Hawthorne effects by assigning students to a blind condition, where some students receive random or arbitrary learning strategies (as defined by Ausubel, 2010), and some receive legitimate literacy-improvement strategies as defined in the content-literacy literature (e.g., Daniels & Zemelman, 2014). This would help differentiate the actual effects of explicit literacy instruction from a placebo.

Fourth, studies are needed to truly isolate the effects of reading music notation and textual notation. The current study featured literacy-enriched strategies that applied the strategies to both music reading and textual reading. The textual reading opportunities were mostly presented as descriptions and clarifications printed in the method book. The concert information handout also presented opportunities for the application of textual reading strategies. Isolating the strategies to only music reading would be possible by simply using a method book written in a language not understood by most of the students in the sample (e.g., German, French, Japanese, or Russian). In such a setting, all literacy strategies would apply solely to the musical notation, and all instructions and clarifications would be given verbally, thereby isolating and identifying only the effect of the textual literacy strategies on music notation.

Fifth, although the current study focused on the reciprocal relationship between music literacy and textual literacy in band, similar investigations are needed to explore the effects of literacy-enriched instruction in related disciplines such as orchestra, chorus, general music, and emerging (non-traditional) ensembles. The design may also be applied to other content areas
such as math, science, social studies, or even art or physical education, which is similar to Baker’s (2011) study, which investigated the relationship among all of the arts, literacy, and math achievement. Most of the content-literacy texts used in the current study are intended to be used in science, social science, and math applications, so an investigation into the efficacy of literacy strategies in those contexts would also be appropriate (Daniels & Zemelman, 2014; Urquhart & Frazee, 2012; Vacca et al., 2011).

Sixth and last, the current study suffers from several limitations related to the chosen research design. Namely, a quantitative design fails to provide rich descriptions of student and teacher experiences. Mixed-method designs and qualitative designs would help provide detailed accounts of the participant experiences while seeking a context for applying literacy strategies more effectively.

Qualitative approaches include collecting case studies of students participating in a literacy-enriched band, naturalistic or ethnographic descriptions of literacy-enrichment in the field, or historical or documentary research related to the development of parallel strategies and their application to many classes. These approaches could also investigate the difference between intended literacy strategies and actual literacy strategies as observed by both teachers and students. They could investigate what materials were helpful during instruction, how time was used, and how age and musical experience play a role in the developmental expectations of the instructor. Student and teacher reflections could be coded for patterns and themes. Multiple method books could be analyzed for literacy-enriched content. The addition of a survey could help introduce a mixed-methods approach to the inquiry, followed by field work and follow-up interviews. In essence, the possibilities for mixed-methods or qualitative investigation are plentiful and would provide an enormous resource for the field of music education.
Conclusion

Music literacy is one of many paths toward a holistic understanding of literacy achievement, as textual literacy is one of many paths toward understanding society as a whole. The world cannot be summarized or experienced solely in a written form. However, many band teachers approach music as if the written form is paramount. Band teachers must broaden their understanding of music literacy and be amenable to the musical and academic benefits of incorporating more student-centered learning strategies. These benefits allow beginning band students to travel a more efficient path, though not the only path, toward an ultimate goal of aesthetically understanding and enjoying of music.

Ironically, it is through the explicit instruction of literacy techniques in music that beginning band teachers are growing beyond a traditional environment that only values reading and performance. The content area literacy texts and influential music pedagogues such as Kodaly, Orff, and Suzuki, deliver the same message: simply reading and rereading material doesn’t work. Students need steps, strategies, and skills. They need time to interact with each other, process the content, apply the content, and reflect on the content. Yet despite their sage advice, little has changed within the prototypical beginning band paradigm. Perhaps this is due to the content of the method books, or perhaps this is simply due to tradition. But band directors are losing an opportunity to connect with students on a deeper level if they only address one type of literacy in an autocratic, teacher-centered environment. When attrition is a continuous concern, this is an area that lies within the teacher’s control and could make a significant impact for retaining students and creating life-long participants in performance music.
The findings of this study reinforce the idea that music should not be studied in isolation. The inclusion of ELA strategies does not diminish musicality or performance-readiness, but it could make music learning more meaningful. In turn, music study does not diminish the development of literacy skills, despite the time and resources required of music students. Under some situations, band study may cause literacy growth. This claim is not made lightly, and may only be supported by a strong theoretical foundation and an explicit description of operational constructs.

Many questions have yet to be answered, and a combination of qualitative, experimental, and non-experimental methods will be required to answer them. The lack of corroborating studies is unfortunate, but the details of the design and implementation were explicitly detailed in this study for the benefit of replication. It is hoped that a program of related studies may be developed to investigate these research questions among more populations. Causation will continuously be questioned until a body of corroborating literature can be developed to confirm the findings through meta-analyses. Such research is paramount to the advancement of the field of music education. As stated in Chapter 2, the enduring consideration of music as a core academic subject largely depends on the reculturalization of society about music’s important artistic and academic benefits. The current study represents a significant step toward identifying and building upon those benefits.
REFERENCES


Andren, K. J. (2010). *An analysis of the concurrent and predictive validity of curriculum based measures (CBM), the Measures of Academic Progress (MAP), and the New England Common Assessment Program (NECAP) for reading* (Doctoral dissertation, University of Southern Maine).


Scientific discourse: Multiliteracy in the classroom. New York, NY: Continuum

The music and literacy connection (2nd ed.). Lanham, MD: Rowman & Littlefield.

Aural skills: At the juncture of research in early reading and music literacy. Music Educators Journal, 99(2), 75-80.


The impact of high-stakes testing on curriculum, funding, instructional time, and student participation in music programs (Doctoral dissertation, University of Florida).


Kemmerer, K. P. (2003). *Relationship between the number of hours spent in general music class and reading skills in kindergarten through grade 3* (Doctoral dissertation, Lehigh University, Bethlehem, PA).


Urquhart, V., & Frazee, D. (2012). *Teaching reading in the content areas: If not me, then who?* Alexandria, VA: ASCD.


APPENDIX A

RECRUITMENT LETTERS
Hey, Band Directors! I am looking for an experienced beginning band teacher who would be willing to host a study investigating the relationship between text literacy and music literacy. You would instruct band one grade prior to your current beginning grade level before or after school. You will be compensated for your time. If you would like more information, please e-mail me. Thank you!

Dave Carroll
Letter Requesting Director Participation

Hello Colleague,

My name is David Carroll, and I am an elementary and junior high band director. I am also in the final stages of my doctoral studies at NIU.

I am writing to you to ask for your help. My dissertation seeks to investigate the relationship between music literacy and text literacy, with a theoretical framework that they are mutually beneficial. My study will be a quantitative, randomized experimental design.

After performing an extensive review of the existing research literature, I am proposing a study design and methodology that requires instruction from a relatively experienced band teacher. In my opinion, the results of this much-needed study may help define the role of music as a means for improving literacy achievement. It will also help music teachers rethink how they teach literacy skills in general.

My design calls for a very specific set of circumstances. Most school districts, including my own, do not qualify. The district must begin band instruction in grade five or later, it must have a diverse student population, and it must already administer the NWEA MAP test. I believe that your district may qualify.

I am interested in running an after-school band program in fourth or fifth grade at one of the elementary schools in your district. It would be taught by you or by a volunteer band director already employed by your district. The program would be absolutely free for your students. The director would be paid a stipend of approximately $30/hour. Additionally, the hosting school would be compensated for the use of space and facilities. I will provide teaching materials and use whatever method book your district already uses. Funding is currently arriving from various grant sources and foundations.

Students will be assigned randomly to a band group or control group to investigate differences in literacy achievement. Once data are collected, students in the control group will get to participate in band with the same number of instructional minutes as the band group so that no student who wants band is denied instruction.

As an elementary band director myself, I am aware of the logistics of such a program, and I am willing to work within your existing curriculum so that nothing we do detracts from your current program. In fact, this study may even benefit your band program while also benefitting the next generation of band and orchestra students. This is a major undertaking, and I am hoping that you might consider meeting with me to discuss how this might work in your district.

I look forward to your response. Thank you so much!

-Dave Carroll
APPENDIX B

PARENT AND STUDENT NOTIFICATION OF RESEARCH STUDY
Dear Parents and Students,

Students are now invited to be a part of a special research study being conducted by David Carroll, a doctoral student at Northern Illinois University. This study will investigate whether playing a band instrument helps reading skills and whether certain reading skills help students play their band instrument better.

Band will be offered AT NO COST as an after-school activity to SOME students this fall and the REMAINING students in the spring. By the end of the study, everyone who wants band instruction will have received the same number of lessons and rehearsals with the same benefits of instruction.

There is a possibility that a student will be assigned to be in a non-band group in the fall and NOT want band instruction in the spring. That is okay, and these students are still needed to determine differences in reading scores. But all students will have the chance to learn a band instrument in fourth-grade if they choose.

Students who are part of the study will be randomly assigned to one of three groups:

<table>
<thead>
<tr>
<th>Winter</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No band in the winter</td>
<td>Traditional band for 14 classes, after school, 1 day/week</td>
<td>Literacy-enriched band for 14 classes after school, 1 day/week</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Concert in April</td>
<td>Concert in April</td>
</tr>
<tr>
<td>Spring</td>
<td>Option to start in band, 14 classes, 2 days/week, using district default model. Concert in June</td>
<td>Option to continue next year and/or join summer band.</td>
<td>Option to continue next year and/or join summer band.</td>
</tr>
</tbody>
</table>

Students in all three groups will receive instruction, materials, and instrument rental at no cost. Band will be taught by an in-district band instructor, using the same book already used by your district. Band students will enter fifth grade as experienced musicians!

All students at your school in fourth-grade are eligible for this study, regardless of ability, socio-economic status, race/ethnicity, age, or gender. Study participation is optional, and
students and parents may withdraw at any time, but parents/students who withdraw will need to return the instrument if it has been provided by the researcher.

**Study Communication and Group Assignment Notification**
Once the consent and assent forms are collected, students will be assigned to their study group. Assignments and class schedules will be communicated via email, with an option for hard copies if the parent indicates such on the survey. Fall classes will begin during the week of September 19 and continue every week until the week of January 9, excluding the week of Thanksgiving and over winter break.

**Selecting a Suitable Instrument**
Your school will be holding an instrument selection clinic for all interested students, regardless of study participation. During this clinic, expert teachers will work one-on-one with students and help them determine which instrument would be the best fit. This is the same process that students would normally undertake in preparation for fifth-grade band, except it will be offered earlier in the school year for the purposes of this study.

**Obtaining an Instrument**
If you wish the researcher to obtain the instrument for you, a form will be provided by the instrument rental retailer. The researcher will cover rental costs during the term of the study, although the parents and student will still be expected to follow the terms of the rental contract regarding appropriate care of the instrument.

The instrument will be delivered to the site and ready to take home at the first lesson. Once the study is over, the instrument should either be returned to the school or you will have an option of continuing rental by assuming rental payments subject to the terms of the instrument rental retailer contract.

If you already own your instrument, please have it checked by a reputable shop to make sure it is in good playing condition. Please ask your child to bring it to the first lesson.

You may also rent your own instrument from any vendor you choose. If you choose this option, you may continue to rent to the instrument for as long as you wish, subject to the terms of the rental retailer contract. Please have your child bring the instrument to the first lesson.

**What to do now?**
- Please consider participating, especially if you or your child would not have ordinarily chosen to be in band. Students can still wait and join band next year if they choose.

- Please be flexible with after-school commitments this fall. Once the groups and schedules are set, they cannot be changed, except if a child or parent(s)
withdraws from the study. The fall session will run from September through January. The spring session will run from January until March.

- Please consider helping other students with transportation (carpooling), as we are trying to eliminate conflicts that may keep students from fully participating in this study.

- Remember that if your child is not selected for the band-instruction group, he or she may still learn an instrument in the spring session instead. The fall and spring sessions will receive equal instruction.

- If you have questions related to the study, please contact the researcher. If you wish further information regarding your rights or your child's rights as a research subject, you may contact the Office of Research Compliance at Northern Illinois University.

Thank you!!

Sincerely,

David Carroll
Northern Illinois University

If you agree to the study, please return the following items:

- Parental Consent Form
- Student Assent (Permission) Form
- Parent Demographic Survey
APPENDIX C

CONSENT AND ASSENT TO PARTICIPATE IN A RESEARCH STUDY
Your child is invited to participate in a research study titled “An Investigation into the Relationship between Text Literacy and Music Literacy among Beginning Band Students” being conducted by David Carroll, a graduate student at Northern Illinois University. The purpose of this study is to determine how music and text literacy strategies may be mutually beneficial.

Study Design

Your child's/ward's participation in this study will last 14 weeks in the fall and an optional 10 weeks in the spring. Parents will be asked to complete the attached demographic information form. Students will then be randomly assigned to one of three groups.

- Students assigned to the first group will not receive band instruction in the fall but will have the option of receiving the same number of lessons and rehearsals in the spring using a hybrid of conventional and literacy-enriched instructional techniques.
- Students assigned to the second group will receive conventional (performance-based) band instruction in the fall and will have the option of continuing in the spring.
- Students assigned to the third group will receive literacy-enriched band instruction in the fall and will have the option of continuing in the spring.

Depending on group assignment, your child may be asked to attend one music lesson and rehearsal each week, culminating in a performance at the end of the initial 14-week fall session and an additional, optional performance at the end of the 10-week spring session.

We will be collecting the MAP reading scores for all participants to check for differences among the three groups. Band students will also take a short music reading test. All scores will be kept confidential.

Risks and Benefits

There are no foreseeable risks and/or discomforts your child could potentially experience during this study. The instruction, materials, books, and instrument rentals will be provided free of charge for all participants during the length of this study. Parents have the option of keeping the materials and/or continuing their instrument rental beyond the term of the study if they so choose, subject to the terms of the retail instrument rental provider of the parent’s choice.

You or your child may decide to withdraw from this study at any time. However, if the instrument was supplied to the student by the researcher, the instrument will need to be returned at the time of discontinuation.

The benefit your child may personally receive from participating in this study is research-based literacy instruction in a band setting at no cost, possibly resulting in increased literacy skills and higher test scores. Successful strategies may potentially transform how beginning band is taught nationwide.

In the event of a research-related medical emergency or if your child should experience an adverse reaction, please immediately contact (name and telephone number of person to contact)__________________________.
Although Northern Illinois policy does not provide for compensation for treatment of any injuries that may result from participation in research activities, this should not be construed as a waiver of any legal rights or redress you or your child might have as a result of participation in this study.

**Reporting of Findings**

Information obtained during this study may be published in scientific journals or presented at scientific meetings, but any information that could identify your child will be kept strictly confidential. Your child’s/ward’s name or other personally identifiable information will not be reported, and all data relating to this study will be kept on a non-networked computer.

**Choice of Participation**

Participation in this study is voluntary. Your decision whether or not to allow your child as well as his or her assent to participate will not negatively affect you or your child. Your child will be asked to indicate individual assent to be involved immediately prior to participation and will be free to withdraw from participation at any time without penalty or prejudice.

Any questions about the study should be addressed to David Carroll. If you wish further information regarding your rights or your child's rights as a research subject, you may contact the Office of Research Compliance at Northern Illinois University.

Please check the appropriate box, sign the form below, and return it to your school.

I ☐ do ☐ do not agree to allow my child to participate in this research study and acknowledge that I have received a copy of this consent form.

I ☐ do ☐ do not agree to allow my child’s instrumental performance to be audio recorded.

____________________________________________ _________________
Signature of Parent/Guardian   Date

If you agree, please also return the • Student assent (permission) form
• Audio recording consent form
• Parent demographic survey
Adult Consent to Participate in a Research Study

You are invited to participate in a research study titled “An Investigation into the Relationship between Text Literacy and Music Literacy among Beginning Band Students” being conducted by David Carroll, a graduate student at Northern Illinois University. The purpose of this study is to determine how music and text literacy strategies may be mutually beneficial.

Study Design

You will be audio recorded during a meeting concerning text reading strategies. Specifically, you will be asked what explicit reading strategies you currently use in your classroom, how you use them, and what time of year you generally introduce them to your students. The recording will be transcribed in order to assure accuracy. You will have an opportunity to approve, deny, or clarify statements that you make before the transcript is printed.

Your name and school will be kept confidential in the reporting and presentation of this study.

Risks and Benefits

There are no foreseeable risks and/or discomforts you could potentially experience during this study. You may benefit from a rich discussion about how and when you address explicit reading strategies in your classroom. You will not be asked to change your current practice unless the change would have occurred independently from this study.

Although Northern Illinois policy does not provide for compensation for treatment of any injuries that may result from participation in research activities, this should not be construed as a waiver of any legal rights or redress you or your child might have as a result of participation in this study.

Choice of Participation

Participation in this meeting is voluntary. Any questions about the study should be addressed to David Carroll. If you wish further information regarding your rights or your child's rights as a research subject, you may contact the Office of Research Compliance at Northern Illinois University at (815) 753-8588.

Please check the appropriate box, sign the form below, and return it to your school.

I ☐ do ☐ do not agree to participate in this research study and acknowledge that I have received a copy of this consent form.

______________________________________________  __________________
Signature of Adult                              Date
Dear student,

I am doing a study on whether playing in the band helps you read better. If you agree to be in my study, I can investigate the answer by putting you into one of three groups:

- **The first group** will not be in the band in the fall, but could choose to play in the band in the spring. Don’t worry, you can get the same number of lessons (18) in the spring if you want them.

- **The second group** will be in a “regular” band in the fall, which will meet after school one day every week for 14 lessons. We’ll have a concert in January. If you want to continue, we’ll have four more lessons in the spring so you can perform with the first group!

- **The third group** will be in a reading-enriched band in the fall, which will also meet after school one day every week for 14 lessons. We’ll also have a concert in January, and you can also continue in the spring with four more lessons so you can perform with the first group!

Some students may not be assigned to a band group and may choose not to be in the band in the spring. That is okay, but I still would like your permission to see how you read with or without band instruction. You can ask questions about the study at any time. If you decide not to finish the study, you may stop at any time.

If you want to be in the study, please sign this paper. Don’t sign the paper if you do not want to be in the study. Being in the study is up to you, and no one will be upset if you don’t sign the paper or if you change your mind later.

Thank you for considering to be in my study!

Your signature________________________________________ Date______________

Print your name_________________________________________
APPENDIX D

CONFIDENTIAL DEMOGRAPHIC SURVEY
Confidential Demographic Information Form

Please return this form to your teacher in a sealed envelope. Required fields are marked with an *

1. Student’s Name*_______________________________________________________

2. Parent/Guardian’s Name(s)*__________________________________________

3. Parent/Guardian’s Phone Number(s)*_____________________mobile_________________

4. Parent/Guardian’s Email address ____________________________________________

Communication will be mainly sent through email. Do you prefer hard copies?____

5. Age*______ 6. Gender_______ 7. 4th grade class teacher_____________

8. Not all students will be placed in a band-instruction group. If your student IS randomly selected to participate in the band-instruction program, what are the top three instruments that might interest your child? Please number the choices with a 1, 2, and 3.

____Flute ____Trombone
____Oboe ____Baritone
____Clarinet ____Tuba
____Trumpet ____Percussion
____French Horn ____I am not sure

9. If your child is placed in a band-instruction group, please indicate how you would like your instrument provided*: 

☐ We already own an instrument
☐ We will use an instrument provided by the researcher at no cost
☐ We will rent an instrument on our own
☐ Other___________________________________________________________

10. If your child is NOT placed in a band instruction group in the winter, would your child like to participate in the optional band session in the spring?* Spring students will receive the same amount of instruction as winter students. ☐ yes ☐ no

11. Please list any days of the week that will NOT work for after-school lessons for you or your family. Indicate if the conflicts happen in the winter, spring, or both.

________________________________________________________

12. Are you able to help carpool if needed? ________
13. Has your child taken private music lessons outside of school? □ No □ Yes: Instrument________
               How long? ________

14. Which of the following best describes your race or ethnic background?
   □ Asian
   □ Black/African American
   □ White/Caucasian
   □ Hispanic/Latino/a (may be any race)
   □ Native American
   □ Other (Please specify)_____________________
   □ Prefer not to say

15. Is your child eligible to receive free or reduced-price lunches at school under the National
    School Lunch Program?
   □ Free
   □ Reduced Price
   □ Prefer not to say
   □ Neither

16. Have you lived at your current address for more than two years?
   □ Yes
   □ No
   □ Prefer not to say

17. Can you best describe your home environment?
   □ 2 parents/guardians
   □ Single parent/guardian
   □ Prefer not to say

Thank you for your participation!!
APPENDIX E

STUDY BUDGET AND EXPENSES
Study Budget and Expenses

<table>
<thead>
<tr>
<th>Description</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>English to Spanish translation of study documents, consent, and assent forms</td>
<td>$100.00</td>
</tr>
<tr>
<td>30 used tier-one (high quality) instruments*</td>
<td>$5,136.18</td>
</tr>
<tr>
<td>43 new method books, folders, and pencils from local music vendor</td>
<td></td>
</tr>
<tr>
<td>Copies, envelopes</td>
<td>$150.12</td>
</tr>
<tr>
<td>Instructor Stipend</td>
<td>$3,330.00</td>
</tr>
<tr>
<td>Total</td>
<td>$8,716.30</td>
</tr>
</tbody>
</table>

*Some participants used instruments were donated, loaned, or school-owned, which reduced the number of instruments purchased through the vendor

Fundraising

Prior to fundraising, a fiscal agent was established to collect and disperse costs associated with the study. The agent was the Naperville Unit Education Association, a 501(c)(5) organization under the Internal Revenue Service Code, which lawfully reported taxable earnings for the instructor’s stipend. The stipend and the instruments were the two largest costs of the study and were necessary to provide participation in the study at no cost to the students.

Funding came in the form of grants and donations. Major grants were awarded from the Louise M. Berman Foundation and the National Education Association (NEA) Foundation. Individual donations were collected through a Go Fund Me website. All grantees and donors gave irrespective of the study results. That is, notifications were clear that there were no guarantees for specific results or outcomes because of the status of a donation or award. Student incentives (coupons) were donated by local restaurants. Efforts were made to assure the ethical and transparent collection and distribution of funds during the implementation of this study.

This appendix is provided to assist in the future replication of this study.
APPENDIX F

INSTRUCTOR GUIDELINES AND LESSON PLANS TO DIFFERENTIATE BETWEEN EXPERIMENTAL TREATMENTS
Instructor Guidelines to Differentiate Experimental Treatments

Note: Instructor training occurred three times prior to the initiation of instruction. Training included the theory and examples of conventional and literacy-enriched strategies. The instructor kept a written log of weekly lesson plans and strategies used.

Students: 4th Grade band students with no prior band experience

Timeline: 14 weeks, with one small group lesson and one full ensemble rehearsal each week. Each lesson is 30 minutes in duration. Each full band rehearsal is 50 minutes in duration.

Materials: Students used Measures of Success (Sheldon, et al., 2009) for flute, clarinet, trumpet, trombone, and percussion.

Objective: By the end of this unit, students will able to describe and demonstrate fundamental skills on their chosen instrument, including
   • appropriate instrument care and maintenance for a first-year player
   • proper posture and playing position
   • basic music notation, including the staff, bar lines, quarter note and half note/rest graphemes
   • seven pitches utilizing a range appropriate for a first-year player (notes vary by instrument)
   • full ensemble rehearsal technique and etiquette
   • a demonstration/performance for parents and public attendance

Additional group 1 objective: Students will use conventional rehearsal and lesson techniques to accomplish the above objectives. These include clapping, singing, rote teaching and modeling, trial and error, chanting, repetition, and chunking (dividing material into its smaller component parts), trial and error practicing, teacher/peer modeling, verbalizing note names, verbalizing fingerings, audiation, and repetitive drill.

Students may complete any non-playing activity indicated in the method book. Students may independently initiate literacy strategies described below, but the instructor will not initiate those activities in group 1.

One recurring activity used in conventional instruction is known as the Four Step Practice System. This consists of the following four steps:
   1. clapping the rhythm
   2. clapping while saying the letter names
   3. moving the fingers and saying the letter names while another student or teacher plays
   4. student plays

Additional group 2 objective: Students will use literacy-based strategies to accomplish the above objectives. At least one strategy is to be used at every lesson and at every full ensemble
rehearsal. Group 2 students will receive more time to write in the method book. Writing activities will be purposeful and rehearsed, especially in the early stages. After week seven (the halfway mark) the gradual release of responsibility will begin to help ensure that students use the strategies independently with fidelity.

The method book introduces new concepts in textual form via short vignettes, or blue boxes. For example, line 1.19 features a blue box that states, “Style and Form: Duet. A duet has two different parts performed simultaneously by two individuals or groups” (Sheldon et al., 2010, p. 8). Group 1 members will receive a verbal introduction to the content. Group 2 students will utilize the turn-and-talk, summarization, vocabulary predictions, annotation strategies, or other listed strategy to help them recall and understand the content.

The strategies are summarized below:
This chart was developed after meeting with a fourth-grade representative teacher about literacy strategies. These strategies were each to be used at least once in each lesson and full band during the study. Strategies may be used more than once.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Done in 4th grade ELA classroom?</th>
<th>How might this transfer in music?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation:</td>
<td></td>
<td>Seldom</td>
</tr>
<tr>
<td></td>
<td>Students explicitly stop reading/playing to jot notes and reminders in the margins of their book. Annotations specifically refer to writing words and phrases that help students remember a central concept or unknown passage. The teacher models the process using a projected example, and then asks students to demonstrate their own ability to annotate in their own book. Annotations should not be confused with coding, during which students write symbols, small pictures, or circles for the same purpose.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Limited usage in beginning level music class.</td>
<td></td>
</tr>
<tr>
<td>Coding:</td>
<td>Frequently, especially in &quot;close readings,&quot; which are short, high interest readings.</td>
<td>Model how you write codes into music. Use the same codes as 4th grade…!=important, ?=question, star=WOW. 4th grade has a guide they use.</td>
</tr>
<tr>
<td></td>
<td>Seldom, but yes.</td>
<td>Show pictures of themes, instruments, diagrams, and new symbols. How does the form follow the function of the notation?</td>
</tr>
<tr>
<td>Frontloading with Images:</td>
<td>The teacher prepares students for learning by using photos or drawings to establish the context, problem, or process of the lesson’s topic. The images are projected on a screen and the teacher allows time for students to look carefully. This process allows students to make predictions and inferences in a manner that can engage students of all ability levels. In band, images can help provide a memorable context for the music to be practiced, especially since the chosen method book features several short vignettes about composers and time periods.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Seldom</td>
<td></td>
</tr>
<tr>
<td>R-D-W:</td>
<td>This stands for Read-Draw-Write. This strategy was described by Debbie Grawn (2015) as a strategy originally developed for the Eureka Math system. The system involves reading and rereading the problem, then drawing the situation represented from the given information, and finally writing conclusions from the drawings in any textual or numerical</td>
<td></td>
</tr>
</tbody>
</table>

**R-D-W**: This stands for Read-Draw-Write. This strategy was described by Debbie Grawn (2015) as a strategy originally developed for the Eureka Math system. The system involves reading and rereading the problem, then drawing the situation represented from the given information, and finally writing conclusions from the drawings in any textual or numerical
format. This helps students use their drawings to understand the problem. It also helps teachers see where students may be misinterpreting the problem by illustrating their logic and thought process.

| Seldom, but yes. | May be used to illustrate phrase structure or melodic line. May also be used to graphically plan a problem-solving strategy for practicing difficult material. |

**Post-it Response Notes:** Students use post-it response notes while playing music to jot questions, surprises, and reminders. This helps students remember and practice unknown material at a later date. Teachers can specify certain themes for the day or allow students to jot information as they see fit. The post-it notes can be collected to create an archive of student growth.

| Seldom now, but increasing in the future. | Use post-its for new notes, vocabulary, or a difficult sequence for later practice. Emphasize that this can be done in reading, too. |

**Read Aloud:** The teacher or a student models the performance of a short passage of interesting music. Modeling includes fluency and inflection that are stylistically appropriate. Material should not be chosen from the textbook, and should represent various levels of reading/playing ability. Different than rote teaching, read alouds typically feature example music (or text) that does not include student material.

| Frequently | Play for students. But to make the reading connection, help tell them what to listen for. |

**Summarization** (usually done in writing)

| Done in different ways as described throughout this document. | See strategies below. |

**Think aloud:** The teacher demonstrates her thought process out loud while reading an unfamiliar passage of music or text. This process allows students to see how the teacher handles unknown or difficult material. Rather than overlooking unknown symbols, notes, or words, the teacher demonstrates how to go about looking up the information. After modeling the process, students try it in small groups. This is a great way to demonstrate how and when to use practice strategies such as altering the tempo, chunking, and repetition.

| Frequently | Explain your thought process while you model sight reading. |
**Turn and Talk:** Students discuss important concepts introduced by the teacher for one to two minutes. This is followed by a brief report out from a selected sample of students. This is also referred to as “think-pair-share,” and can be used many times within a single class period. This strategy helps engage students who would otherwise be disengaged during class or confused about the topic. Plus, it allows the teacher to evaluate students' readiness and understanding quickly.

Frequently

This can be done during lessons easily. But to get better in reading class, give more direction. We're not just looking for your first thought, but your BEST thought.

**Vocabulary Predications:** Students are given a list of several potentially unknown words, which they then attempt to categorize by predicting their meaning. The teacher provides the list of words, and students attempt to sort them in small groups. Students then attempt to predict a common theme for each group of words. The strategy uses collaboration and prior knowledge to help students master new vocabulary.

Frequently, especially tier 2 words.

We don't do this nearly enough in music. But prediction can help show the similarities between Italian and Spanish, especially.

### Other Strategies (not to be used in this study, but possible for future studies)

<table>
<thead>
<tr>
<th>Strategy Name</th>
<th>Frequency in 4th grade, per interview</th>
<th>Application to band/music instruction.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-reading quiz</td>
<td>Often</td>
<td>Can be used for major concepts or when multiple vocabulary words are introduced.</td>
</tr>
<tr>
<td>Dramatic role play</td>
<td>Often</td>
<td>Can perhaps be used as a means to associate pitches with fingerings (storytelling, etc.)</td>
</tr>
<tr>
<td>Exit/Admin Slips</td>
<td>Moderately, but usually informal</td>
<td>Can be used to quickly judge understanding or prior knowledge</td>
</tr>
<tr>
<td>K-W-L (Know, want to know, learned) discussion</td>
<td>Moderately</td>
<td>Can be used for especially major concepts.</td>
</tr>
<tr>
<td>Multi-column notes</td>
<td>Frequently, but they call it “double entry journals or T-charts”</td>
<td>Use when comparing two ideas. Maybe try this later, but think about it.</td>
</tr>
<tr>
<td>Partner reading</td>
<td>Frequently</td>
<td>Like the turn and talk, but more active.</td>
</tr>
<tr>
<td>Activity</td>
<td>Frequency and Notes</td>
<td>Additional Information</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Password</td>
<td>Seldom, but worth trying</td>
<td>Definitely worth a try as students become more familiar with symbols and terms. Students may do this at home with other subjects, too, particularly reading.</td>
</tr>
<tr>
<td>Sketching</td>
<td>In math only, and even then it is highly regulated, not creative.</td>
<td>Can be used when teaching D.S. al coda, repeats, etc.</td>
</tr>
<tr>
<td>Tweet the text</td>
<td>Never, but may try in spring</td>
<td>Nice summary. Encourage kids to do this independently.</td>
</tr>
<tr>
<td>Vocabulary Tree</td>
<td>Done, but not in &quot;tree&quot; form</td>
<td>Looking at patterns. Maybe best done with rhythms. Might have to be creative here to make the association to reading class.</td>
</tr>
<tr>
<td>Written conversation</td>
<td>Never, but worth trying</td>
<td>Worth trying as a summary.</td>
</tr>
<tr>
<td>Where do you stand?</td>
<td>Seldom, but worth trying (in Jan.)</td>
<td>Maybe not applicable in beginning stages of learning an instrument.</td>
</tr>
<tr>
<td>Word wall</td>
<td>Frequently, but done in binders instead of on a wall. No room on the wall.</td>
<td>Maybe have students add vocabulary and symbols to their own binders?</td>
</tr>
</tbody>
</table>
Instructor-Generated Log of Lesson Sequence

Group 1 (conventional instruction) lesson sequence.

<table>
<thead>
<tr>
<th>Lesson and Full Band Number</th>
<th>Group 1 Lines&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Group 1 Knowledge and Skills Focus (Conventional Instruction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Instrument assembly and hand position</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fundamental tone production (buzzing, blowing)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lesson organization and procedures</td>
</tr>
<tr>
<td>2</td>
<td>1.1-1.3</td>
<td>Winds: Staff notation, Concert “D” and “C,” Quarter notes/rests</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion (Starting on Snare): Staff notation, Stickings,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quarter notes/rests</td>
</tr>
<tr>
<td>3</td>
<td>1.4-1.6</td>
<td>Winds: Alternating between 2 pitches w/o rests between</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Sticking in groups of 2 and 4</td>
</tr>
<tr>
<td>4</td>
<td>1.7-1.10</td>
<td>Winds: Concert “Bb,” Alternating between 3 pitches</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Alternate sticking, Reduced pattern predictability</td>
</tr>
<tr>
<td>5</td>
<td>1.11-1.12</td>
<td>Winds: Half notes, whole notes, fermata, time signature</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Introduce bass drum, right-hand lead, half notes,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>whole notes, fermata, time signature, rhythmic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>independence between snare and bass</td>
</tr>
<tr>
<td>6</td>
<td>1.13-1.15</td>
<td>Winds: <em>Hot Cross Buns</em>, draw measure lines, write in note</td>
</tr>
<tr>
<td></td>
<td></td>
<td>names, repeats</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Hot Cross Buns, draw measure lines, write in note</td>
</tr>
<tr>
<td></td>
<td></td>
<td>names, repeats, exercises duplicated on bells</td>
</tr>
<tr>
<td>7</td>
<td>1.16-1.18</td>
<td>Winds: <em>Mary Had a Little Lamb</em>, <em>Au Claire de la Lune</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: <em>Mary Had a Little Lamb</em>, <em>Au Claire de la Lune</em>,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>muffling</td>
</tr>
<tr>
<td>8</td>
<td>1.19-1.20</td>
<td>Winds: Duet, Rhythmic independence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Duet</td>
</tr>
<tr>
<td>9</td>
<td>1.21-122</td>
<td>Winds: Concert “Eb,” alternating between 4 notes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Multiple bounce stroke</td>
</tr>
<tr>
<td>10</td>
<td>1.23-1.26</td>
<td>Winds: Concert “F,” breath mark, drawing notation elements</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(clef, time signature, final bar line)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Continued development of the multiple bounce stroke,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>drawing notation elements (clef, time signature, final bar line)</td>
</tr>
<tr>
<td>11</td>
<td>1.27-2.1</td>
<td>Winds: <em>Good King Wenceslas</em>, <em>Go Tell Aunt Rhody</em>, <em>Ode to Joy</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: <em>Good King Wenceslas</em>, <em>Go Tell Aunt Rhody</em>, <em>Ode to Joy</em></td>
</tr>
<tr>
<td>12</td>
<td>2.2-2.3</td>
<td>Winds: Concert “G,” <em>Old McDonald</em>, alternating between 6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>notes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Percussion: Flam, <em>Old McDonald</em></td>
</tr>
<tr>
<td>13</td>
<td>2.4-2.5</td>
<td>Winds: Concert “A,” Phrasing, Alternating between 7 notes</td>
</tr>
</tbody>
</table>
### Percussion: Alternating between flams and multiple bounce strokes, long roll (fermata)

| 14 | 2.6 | Winds: Key Signature  
Percussion: Continued flam development |

a. Lines refer to exercises from the *Measures of Success* (Sheldon, et al., 2009) method book

#### Group 2 (literacy-enriched) lesson sequence and strategy application.

<table>
<thead>
<tr>
<th>Lesson and Full Band Number</th>
<th>Group 2 Lines(^a)</th>
<th>Group 2 addressed the same skills and knowledge as group 1, plus the following additional strategies (Literacy-enriched Instruction)</th>
</tr>
</thead>
</table>
| 1                           |                      | Frontloading with images (Exploring the format and images of the method book.)  
Vocabulary predictions (Predicting articulation, pitch.) |
| 2                           | 1.1-1.4              | Summarization (Summarizing the differences between quarter notes and whole notes. Also, summarizing vocabulary in the blue boxes, including staff, bar lines, and final bar line.) |
| 3                           | 1.5-1.6              | Turn and talk (Associating the name of the songs with their content, e.g., *Up and Down.*) |
| 4                           | 1.7-1.8              | Think aloud (Asking: How is the new note produced? How should it sound? How do you overcome mistakes and misconceptions?) |
| 5                           | 1.9-1.10             | Frontloading with images (Showing students images of clefs and asking them to predict what different clef characteristics mean (the loop in the G clef or the dots on an F clef, for example.)) |
| 6                           | 1.10-1.12            | Coding (Remembering to observe the repeat sign.) |
| 7                           | 1.13-1.16            | Think aloud (Asking: How many beats are in a measure? How do beats influence where bar lines are drawn?) |
| 8                           | 1.17-1.18            | Turn and talk (Asking: What are the differences between two similar songs?) |
| 9                           | 1.19-1.20            | Turn and talk (Asking: How did your partner perform the duet? What advice can you give to help your partner?) |
| 10                          | 1.21-1.22            | Turn and talk (Asking: How is the new note different from the other notes? How does your face (embouchure) change to make the higher notes speak?) |
| 11                          | 1.23-1.25            | R-D-W (Asking: Considering the historical information presented for 1.25, what patterns do you see or how are the events related?) Note: The vignette describes several events occurring at roughly the same time: Stephen Foster’s compositions, Frederic Church’s paintings, the Gettysburg Address, and Charles Dicken’s A Christmas
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 12 | 1.26-1.28 | Vocabulary predictions (Asking: What is the double meaning of the word “scaling?”)  
Think aloud (Exploring the process of conducting.)  
Summarization (Using music to tell a story or express a feeling, work with a partner and perform a story for them. Ask 1 question about partner’s story or feeling.) |
| 13 | 1.29-1.31 | Turn and talk (Discussing the mechanics necessary to create a new note.)  
Post-it response notes (Students read concert handout and develop questions.) |
| 14 | 2.1-2.6 | Annotating (Marking important items on the concert flyer.)  
Frontloading with images (Interpreting the shape and meaning of the fermata.)  
Read aloud (Reading the concert instructions with fluency.)  
Summarization (Concert handout instructions were given.) |

a. Lines refer to exercises from the *Measures of Success* (Sheldon, et al., 2009) method book
APPENDIX G

TRANSCRIPT FROM FOURTH-GRADE COORDINATION MEETING
Transcript from Fourth-grade Coordination Meeting

Notes: This is a transcript of an interview with Mrs. Collins, fourth-grade teacher at the study site. The purpose of the interview was to coordinate the timing and content of ELA strategies with music reading strategies employed in the literacy-enriched band group. As a reference, this interview described reading strategy protocols established in Daniels and Zemelman (2014).

DC: Okay, thank you. And thanks again for being here today. In this packet are some examples of some literacy strategies I just want to go through. We don't have to read this whole thing. I will leave it with you if you want to look at it. [Band teacher] is looking into a copy of this also, and this can jump start our conversation as far as what you do for literacy strategies, how you teach them, and when you teach them, how you approach it and so just to get the conversation going. Okay so the first one: please introduce yourself and describe your professional experience as a literacy instructor.

MC: My name is MC, and I've been a teacher for the past 15 years. I did not start my adult career as a teacher, so I am pretty much, you know, the start of new generation, Common Core, etcetera. Literally, it took me a while. It's not something that is easy to teach. I don't think it's straightforward. It truly depends on each student. Literacy depends on how much background knowledge the student has. It is very structured. If the student doesn't have the structure coming in, you have to start, you know. It is a lot of scaffolding, especially in a dual-language program.

DC: Right, and yours is a dual language class?

MC: Right, yes.

DC: Do you teach word recognition like phonics, or do teach more comprehension; where do you find yourself spending more time?

MC: As a dual language teacher, I spend more time building the foundation for the English application, because up until 3rd grade, students receive more Spanish than English. In 4th grade, this is when the students are supposed to come to a level of 50/50 in the language. When you don't have the foundation of the structure of English, things don’t make sense. In Spanish, things make sense. The “f” is an f, the “b” is a b, and we don’t have “gh” or “ph.” I tell my students this is crazy because this sound can be made by this combination, by this blend, by this blend, and that is something that they have to learn. And I saw [native-English] students start learning those foundational skills in kindergarten. My students, sometimes, the first time that they've seen it is with me which is not right.

DC: Sure, absolutely.

MC: And it shouldn’t be. My students are still writing English in Spanish phonetically, so I do four different ways of word study.

DC: Okay.
MC: I have academic vocabulary.

DC: Okay

MC: I have spelling. I have word sorts and (what is the other one…) the other one escapes me…oh, and language, like suffixes and prefixes. I believe that is it.

DC: Okay. So can you tell me an example of how each one of those might look as you talk about it in your class?

4:20 MC: When we do academic vocabulary, in many instances, we front-load the vocabulary so they can have the background knowledge of the meaning before they're starting a specific lesson. And that is in content areas like math, social studies, science.

DC: Okay

MC: Word sorts is language structure, is blends of how the letter words and sounds work. It can be also some prefixes and suffixes, homophones, all those. But it starts with long “A,” short “A”, the long and short vowels, and then the combination of long and short vowels, and how different sounds are made by a combination of different consonants and vowels. So the students learn what is a consonant-vowel, silent E, so when they learn the pattern, the idea behind it is if I learn how to spell “right,” I know how to spell “night,” and “fight,” and “light.” And if I understand that a long “A” with a silent “E” is “mat” vs. “mate,” that silent “E” allocates meaning.

DC: So kids are sorting words by patterns…long “A,” short “A,” silent “E.”

7:35 MC: Yeah, it depends on their benchmark. We benchmark them, and then they each have a group. Because most people go by word patterns, but we do have emergent spellers, so we also have to serve those, because if not, they have too many gaps. So I try to make their groups according to their spelling stage.

DC: And so you benchmark them yourself? You don’t have another test that you perform?

MC: No, there is a district test. It is called our spelling inventory. There is an expectation that the kids in our district should be here. We benchmark them. They receive batteries that we have to comply with for the district.

DC: Oh, okay.

8:37 MC: According to that we form our groups. So we have word sorts, we have spelling…the spelling I try to mirror the word patterns that kids are doing. I go through a series based on association. But there are three tiers of words, right? I try to adhere to tier 2 with the spelling because they are commonly used words that don’t necessarily follow any patterns. But I use repetition. What is going to make an imprint? Seeing them. Using them. Applying them.
DC: Right, context.

MC: Exactly. According to Marzano, they have to see that word or use it 35 times before that word is established as part of their lexicon.

DC: Yes. The other one was language acquisition?

MC: I do another word study in which I do the code names. Things that don’t necessarily follow standard rules. Like when we do Spanish to English, we don’t have feminine or masculine articles in English, but we do in Spanish. So, there is a lot of word study in the classroom. And some we start in isolation. I do believe in isolation because we are purposefully teaching a pattern. And some are taught within the close reading, for example.

DC: Okay, this is great information. This is exactly what I’m looking for. This leads right to question number 2. Recently there has been a great deal of attention paid to the explicit instruction of literacy strategies, and this is largely due to Common Core. Common Core has made us look cross-curricularly, and asked us how we can teach literacy better in social science, and math, and PE, and in music. This book [Subjects Matter] evolved out of that culture. I will show you a bunch of the literacy strategies that they recommend, and they're coming at it from a content area standpoint, not a literacy standpoint. So, for a history teacher, for a math teacher, for a science teacher, what are what are some strategies that we can all use? These are exactly the connections that I’m trying to make with music.

13:12 MC: It isn’t necessarily done a lot here in this school but I did a lot in the previous school where I was. In spite of this, I do see it a little with the district, called Text Talk.

DC: Text Doc?

MC: No, Text Talk. You identify specific vocabulary from the text, and you introduce it. It is like a hook. It starts with the first introduction, you know, and it goes from there. It takes the word and then turns an antonym and synonym, making meaningful sentences, and using the word in meaningful ways. But it comes straight from a book and comes from a meaningful reading. It can be a song.

DC: Okay.

14:38 MC: Like for example for my figurative language, I use songs. I use, you know, not just poetry, but we also use songs, and let's see how so-and-so a word is meaningful to them because, “I know, I know that song.” “Did you know that he's a poet?” So we took it from a song, and then with that word, you put the definition in the pictures, and then it has to be also to be tier 2 words. Not tier 3 or tier 1.

DC: A lot of the things that you're talking about are in here [the Subjects Matter book]. The nice thing about this is it doesn’t formalize it because there's no one way to do all of those things. But
I think you'll find that there's a lot of really good ideas in here that you might be able to steal and use.

MC: Believe you, me.

DC: I know when I first read this book, I thought, “Wait a sec. I already kind of do that stuff, but that's so much better!” There are a lot of light bulbs here, so this was a great read for me and this is one of the reasons I went in this direction [with my study]. What I wanted to do is just go through, not in detail, but to go through some of those suggestions and see if this is something you do or not, or if it's something you could think about. You've already talked about how you address these literacy strategies, so I guess looking through here together, the main things that I want to know are whether there are things that you already do, and are there any strategies that you think you might be able to use, and at what point in the year do you think they might be valuable. Then we might be able to do them at the same time over on the other end of the building.

MC: Are these the strategies? (looking at page 90 in Subjects Matter)

DC: These are the strategies with the page numbers. Now roughly, they are arranged with pre-reading first. So basically, we have alerting and some of the contextual strategies you were talking about before you actually read. The middle column indicates things to do during reading—things like coding, and annotating, and writing and marking symbols. And then the last category is after-reading—so things like reflection, turn and talk, draw pictures, do a skit. One that I really like is Tweet the Text…I love it…as a reflection piece at the end of a lesson, have students summarize what they worked on in 140 characters or less.

18:51 MC: Mmhmm. Like an exit ticket. I used to do this activity when I was on Twitter. But I thought you were meaning like online.

DC: For real?

MC: Yeah.

DC: You could, but this is just to get them to keep it concise.

MC: Yeah. If it’s the one on paper, each student had a spot on the bulletin board outside, and each one would write a Tweet as an exit ticket and put it there. But then students who would pass by would take them or rip them. Or we would write with dry erase markers and students would erase them. It didn’t last long. We don’t have room in the classroom for a bulletin board now, but we used to.

DC: Well, that is a great idea. So I'll show you how they organize this [the book]. When they roll these out and describe them…like the Tweet the Text on page 138…so you’ll find the page numbers here. The focus for this one is reading and summarizing [pointing to the book]. This shows you when to use it, and then they describe it. They talk about why to use it, how to use it, and why it works. So for a lot of these, that starts off by saying, well first you need to model it, and then you have them practice it. So as far as this one goes, after they read a passage they
work with a partner to come up with 140 character summaries, which they should refine with the whole group. It'll give some examples. Then it explores why we use it. In this case, we use it because it forces kids to summarize.


DC: So that's one of them, but I love the way they go into more detail. I thought we could totally do that in my band class. So if it's alright with you, what if we went through each page just to get an overview, and ask is this something that you already do is it something that you want to do. And at what time of the year do you think, should kids know how to do it now, or maybe it is something you do after break, or maybe it is something you do after spring break. So the very first strategy is the think aloud.

MC: The think aloud, yes.

DC: So for the think aloud, you put a passage on an overhead, and then you talk to them out loud as you analyze it. You let them know how you are thinking. This word might relate to this. So basically you're bringing them in on your thought process, which is something we do when we read music, too. I'm going to stumble on this word, I need to think ahead. Because once kids know how you see a piece of text or a passage, then they get more comfortable in doing that themselves. Is this something you already do, you could do, or maybe not this year?

24:15 MC: This is one of those must do's and we do it every year.

DC: Okay…

MC: You try to do it more because it is modeling thinking so they can do it too, right?

DC: Yes.

MC: But we already do that, and we basically do it as a release of responsibilities. At the beginning of the year, my books are very thin. We do read-alouds. In the read-aloud, a model what I'm supposed to be doing…with the metacognition, thinking about my thinking, so that's something we did. We tried to do again a gradual release of responsibility so by the end of the school year, I don't do it so much. They are in charge--they're doing it.

DC: Yes, they talked about that a lot and this and this book in the previous chapters. But, yeah that's exactly where we're heading towards. Okay, so that's a good one. So we can turn to 98 and this is reading out loud, so for this, we’re modeling inflection or modeling continuity and fluency, so that kids can learn what it's supposed to sound like in their head while they are reading.

26:19 MC: Since this is a dual language class, more with this class than ever before, I always do the first reading. We always tell them, “You know the first reading is for the flow, is for the gist of it.” This one is for the gist, this one is for the flow. So what you're supposed to be doing is paying attention, so you hear me saying that word, you hear my pacing, listen to how I read it so they can mimic it when they're reading it on their own.
DC: Yeah, I think you can't have comprehension without fluency. And we see that with kids playing rhythm on their instrument, when they take measures one note at a time and it may sound really chunky, we know that they're not they're not understanding the bigger picture. They're not hearing the song. I've had kids play *Twinkle Twinkle Little Star* and they didn't know they're playing *Twinkle Twinkle Little Star* because they weren't fluent enough. Once they got the fluency then the light bulb went off. Then kids say, “I know this song!” “Yeah, you do!” So it's exactly what we're doing too, and that's a great strategy.

MC: That is why we do it, so they know, for modeling.

DC: All right. **Front loading with images**, especially when you are introducing new vocabulary. The goal here is that you put an image up which piques their curiosity. What do you see in this image? And they even put an example on the next page, which is a fascinating social studies example that they use. They took one picture and broke it into 4 smaller pictures. Then they had groups of kids analyze each quadrant, and they each independently talked about what they saw, trying to predict what was happening in each scene. Then they put it all together to hear the whole class come up with totally different because they saw something different altogether.

MC: I like that idea. I like that. I’ve done these a lot in social studies, but never like this. I like to use this as an attention grabber. We call it an attention grabber. And for example, we used to have a snake. When I was starting the civil war, when I was a fifth-grade teacher, we used pictures. But I have never seen it this way. I like breaking it up because now we have speaking and listening. The kids are supposed to be leading the discussion more and this helps. Now is this for vocabulary or for a theme?

DC: It can be either one. According to this, it helps with setting, context, process, problem.

MC: Okay. I like it. I always use it as an attention grabber, but not like this.

DC: I use this if I’m introducing a new piece that has some sort of historical significance to it. Not with beginners, but with a little more advanced students. Not to say that I couldn’t use it with beginners…maybe I should. But there is a lot of benefits to this because kids are so image oriented. They remember images a lot longer than they remember what we say or what we do. Those images last.

MC: I’ve never used that here though.

DC: **KWL**.

MC: Yes, we use that all the time.

DC: Okay. A **pre-reading quiz**.

MC: It depends on what we're doing because it's sort of like another way to do KWL. But it depends on, you know, to maybe check for understanding, but not often.
DC: Okay. The nice thing about their descriptions that date they lay out a couple of different ways to do it, and different types of questions to get different type types of responses depending on your class, and the age, and in the content area.

MC: You know what you don't know is why are you reading this. We just give them what to read.

DC: Yeah, well the nice thing about this is it's not about the right answer. When you're asking me what percentage of world's children do you think are in forced labor, that kind of question spurs a conversation. It's not about being right, it's just about being curious, and that leads into the content that you were going to teach, because they just got a very thought-provoking question.

MC: Oh, yeah. This is very cool. It has a level of engagement. I wonder why didn't I think of this before.

DC: This was such a great book because I did the same thing.

35:39 MC: It makes sense, of course. Because we're so busy, that's why.

DC: Truly, we are. How about dramatic roleplay?

MC: Yeah, we do this a lot. I ask, “Who are my actors today?” We do that a lot.

DC: Good, awesome. Next one…vocabulary prediction.

MC: Like context? What are they trying to do?

DC: Let’s see.

MC: The thing is a lot of the work that comes with the spelling, not that I would give it this name, but we do this. Our vocabulary program does this. They put the word into a sentence, predict what it means based on the prefix. But it gives them the answers, too. It’s more like a pairing, not a…

DC: And this book talks about tier 3 words, not necessarily tier 1 or 2. But in your classroom, this would certainly work with tier 2 words also.

38:46 MC: Yes. But this is nice to do in math, with dividend and divisor for example. They understand they have something to do with division, but which one is which? Which one would you put as the divisor, and which one would you put as the dividend? [Looking ahead.] This is cool.

DC: Yeah, this is partner reading. This is something you do in the context of spelling. My guess is that we’re probably going to get into some that you don’t do quite as much.

MC: That is a constant. We do that a lot.
DC: Okay, but here, they are physically reading it out loud to each other. They aren’t doing an analysis yet. They're just reading to each other, and this strategy. “Between paragraphs, the partners start to discuss and clarify their understanding,” so it starts with reading out loud and then goes to analysis afterward.

MC: Well a lot of our partner reading up until now is in math. It is not in reading class. They are not partner reading yet. Not this group. I've done it before, but this group is not ready for that yet. Now in math, since they're shorter word problems, they do partner reading.

DC: And they discuss strategies for solving the word problem?

MC: And then at the end, if you like the strategy your partner used, then I suggest you continue to use it. You know, things like that.

DC: That is something that is really easy for us to pick up in the band world, because as kids are talking about they can play a passage, they can go back and talk about how they’re going to practice it, or what they need to do better, or how they need to breathe. So that's something that's really easy to transfer over, too. The goal is that if they practice it in both places, then they're going to get really good at it and faster at it. Here's a new one…post it response notes. This is something most people don't necessarily do.

MC: A response to what? To their reading?

DC: So during reading, they’re taking notes on their reactions. But they’re writing them on post-it notes.

MC: And then they have a sheet that they put the notes on?

42:50 DC: Well, there's a couple different ways to do it, but their recommendation is to pause to jot on a small sticky note. “They write the key aspects of the topic, as well as their reactions, questions, or connections. The teacher can specify the kinds of information to focus on. If the books articles are their own, students can leave the notes in the book. If the materials belong to a class library then they can transfer them to their own notebook so that they can review their notes on their own page.” The page can also be used to hold codes or annotations described the next two strategies. My initial thought was that they posted like you have here, but these are more personal.

MC: Yeah, there are different schools and they have different tendencies. Usually, there is a whole school initiative. In my previous school, students had a binder in which the students keep track of their notes. That would be a great idea to implement in this school but we don't do it now. But there is a binder, so when you read all the post-it notes that you used for that day, they are placed on a page and with the title of the book and what page did I just noticed that belongs to. We'll write a reaction to that page, but it's not done in this school.

DC: Okay, and the previous school you taught at was in [this district], right?

MC: Yes.
DC: So that's good. So **annotation**...this is basically kids writing full sentences and words. The opposite would be coding. That's when students write symbols like a question mark or exclamation point or a star.

MC: That we do.

DC: So you probably don't do a whole lot of this because you're using school-owned books.

MC: We do a lot of close reading, and one of the close reading strategies is annotation.

DC: So kids can write directly in their books.

MC: No, well, we don't give them books. We give them short readings, because close readings are short, high-interest readings. The only thing is my class--we're not there yet. We have only done vocabulary, key idea, and supporting details. We haven't done craft and structure yet. So I introduced them little by little, because they are very low. They are very low. So craft and structures we haven't done yet. The next one--they already know how to do main idea and supporting details, and we just finished our first part of summary. Now they know what it is, but I am not saying can do it perfectly. Then after that, we're going to bring craft and structure and that's what we start annotating the text.

DC: So what time of year?

49:05 MC: Right now we should have done it, but since we have such a slow class...in January. We're still working on it. I have to go back and start scaffolding the main idea and supporting details.

DC: Okay, so they gave some examples of annotations.

MC: Those examples are for middle school, maybe grade 6 or so. Our students know the stars, “wow,” and the exclamation point. We have a specific guide that we follow. Use an exclamation mark if this was surprising. Or if I believe this was important, we use a star.

DC: And are those codes by school or by district?

MC: They were not the same at my other school, so I am assuming the guide is only specific by school.

DC: So then **coding** will probably be similar.

MC: Yes, this is what we do.

DC: Just so you know, this is how they are defining annotations and codes. If you see pictures or symbols, they are codes. If you see words, they are annotations. So that’s the difference.

MC: No, we do coding.

DC: **Multi-column notes.**
MC: Yes, we do this. We call them double-entry journals. Is it the theme or is it the main idea? They have to recognize the difference between main idea and theme. What is the point of view? Is it first person or third person? We don’t do second person. There is not a lot of books with second person…I have one book, and that is the only example I have. I show it to them, so they know, but they don’t do anything with it. They pick it up right away. Also, for purpose, they ask, “Is it informative, is it persuasive?” Also, we use them to identify evidence.

DC: Interpretations, reactions, synthesis?

MC: Yes. We call them double entry journals.

DC: They have an example here. Okay…sketching?

MC: Oh, yes, in math.

DC: This may be more of a high school strategy. But if someone has a story problem, it could be useful.

MC: Well in math, they have to illustrate and then solve. They have what they call a “read, draw, write.”

DC: Okay. When we have something similar in band, we’ll call it the same thing. Read-draw-write. When kids have repeat signs, or a DS al coda, and they have to go up and then down over here, they could do a similar sketch or map right here in the music.

54:57 MC: Well in read-draw-write, they do a sketch here in math. It is over the shoulder thinking. And then here, they do the actual math problem. And then at the bottom, they write the answer. There are posters to show the process.

DC: What kinds of things do they draw when they are the first step?

MC: We call it diagrams, or we call it number band. It's all related to that. They can do a number model. This is more Eureka math. It asks more for specific drawings for specific things. Before it was Everyday Math. Before when we did Everyday Math, just make a diagram of what you're thinking.

DC: So that could be anything. It could be a picture…

MC: It could be a stick figure, it depends on the problem. For example, this girl is giving each of her friends 3 cupcakes. So there is three girls with three cupcakes each. You would draw it out. It used to be that way. Now it is very structured with the Eureka math and what you are allowed to draw. It’s more strict and standardized. But it is the only reference that I can give you for sketching. And I know that the little ones drew more, but not so much in 4th grade. I don’t know any fourth-grade teachers that are making drawings of their thinking. I wish…

57:40 DC: Okay. So now we’re getting more towards the after-reading strategies. So turn-and-talk. Is that something we do all the time?
MC: Yeah.

DC: And Tweet the text.

MC: I had completely forgotten about the tweet. At some point I thought this is useful, but yeah.

DC: I like this passage, “If cell phones are available and legal in school, kids can type their summaries on real phones. If not use a laptop or just pretend on paper.”

MC: We do a lot of technology here. We do we share everything through Google Docs. Pretty much everybody has Google Docs I’ve found. They pretty much use their computer for everything.

DC: Do students have student e-mail accounts and do they use them?

58:51 MC: Yes! Completely. For example, my students are going away for their Christmas vacation. We have a Google classroom, and they can access their words and everything. And last year, I had a lot of them and we had a research project due. So some of them worked remotely together. They are very used to working remotely.

DC: And all the fourth grade, fifth grade, and sixth-grade classrooms use this?

MC: They are supposed to. They all have their accounts. And I know that all my kids last year were pros. It’s a lot less paperwork. Share with me. I don’t have to have 30 papers to edit. They are all shared, I make comments on the side, and they fix them.

DC: Now I know you use Infinite Campus. Do you use any other kind of learning management system?

MC: I don’t know. Just Google classroom. It’s just a way to keep everything organized.

DC: Admit slip…let’s see what they say. (Read description). I think a lot of people use exit slips, but not admit slips. But exit slip itself…

1:03:04 MC: Yes, that is something we do. But this class has been difficult to grasp the concept. They aren’t very positive about their conclusions. I like new things because I get bored doing the same things every year. This is nice.

DC: So word wall…that is probably something you do.

1:04:26 MC: Yeah, we have to take them because they are small. So we use them in binders. Once we remove what we are doing at the moment, they stay in the binders for later. I don’t have it on our walls because we don’t have the space. I can’t have them. Since its fourth-grade, I don’t do it.
DC: The band teacher was talking about doing a word wall in the cafeteria. She would have to put it up and take it down. Not just words, but symbols, like fermata, or repeat signs. So what about the graphic organizer?

MC: Constantly, for everything.

DC: It’s not something we do in music a lot. We should.

MC: I don’t even know how it would look.

DC: Well, I don’t know. But when you’re talking conceptually, there are so many parallels. And maybe it’s not about the content, but maybe it’s about the thinking process. And the steps that you involve, like, what are your options for practicing this content? Or how can you organize different paths depending on the problem you’re having. But if we’re all talking the same language, kids will be better at it because they will be used to doing it.

MC: They need it. They need the graphic organizer.

DC: Okay…word meaning graphic organizer.

MC: We have different graphic organizers for multiple meanings, we have them for prefixes and suffixes. There are different ones for word meaning.

DC: On page 150, we have a vocabulary tree, and they put an example on page 152. This is along the lines of what you were talking about with root words like dividend/divisor and all of the different variations that can go off of that tree.

MC: Well, how cool is that.

DC: So it’s all about linking.

MC: We do this, but it doesn’t quite look like this. I like this.

DC: I’m not sure how to apply this to music, but I know there is a way. But you’re already doing this.

MC: The three form will help them a lot. I don’t do this, but I like it.

DC: Okay, so you’re looking at the list-group-label. You talked about this earlier, when you were doing your word sorts. That’s basically it.

MC: And they do have to underline what the words have in common. That’s one of the pre-reading exercises we do.

1:10:50 DC: The clustering and mapping seems to me a lot like what we have been talking about with the drawing and the trees.
MC: But it includes timelines, diagrams, Venn-diagrams. Yes, they are very familiar with these, especially the Venn-diagrams. I like a t-chart myself. But there are a lot of things like mapping that I like.

DC: It’s interesting how many things you do in math that relate to this. I think so much of that is because of your experience. You’re making these experiences for kids because it comes naturally to you. It's best practice. It’s really good.

MC: I wish we had a place that you can grow all year for all the apps that make this so easy. Although, the people that work for Google and all these big social media send their kids to a media free school.

DC: I didn't know that!

MC: I read an article. But I didn't finish it, so I have to finish it. But they send their kids to a media free school. It's more about experiencing and thinking. I do believe that, too. I think in this environment, these kids are so disadvantaged. They really don't have a lot of experiences so I think that technology will help them.

DC: Back when the technology was relatively newer, many teachers were teaching technology for its own sake, and not as a path to something greater. I think we’re evolving past that now. It's just a tool, and if kids aren't thinking deeply and critically, with or without the technology, then it doesn't matter. It's really changed the way we approach teaching—we don't do the gimmicks just to say we did them anymore.

1:14:00 MC: Exactly. That is the SAMR model. Technology is not there just to augment, it is meant to transform. Last year they generated QR codes for their writing, instead of just using the computer to do that. That's just doing the same but on the computer, right? But if you could do more than generate a QR code, then you could be synthesizing and analyzing, which is much better.

DC: Yes. Okay, written conversation. Writing a note to each other in school.

MC: That would be nice. I don’t know how that would go. Legalized note writing. Needless to say, I have never done this. I haven’t seen them.

DC: I can see the potential…

MC: And the debacle.

DC: Yeah, but if you’re already doing Post-it notes for something else, it'd be really easy to take a Post-It note and write a note to your friend. Let them know what you thought about what you just read or how you practiced last night.

MC: Yeah, I like that. But it is basically that form that Eureka math has, which asks, “I like what my partner did because…”

DC: Yeah.
MC: It’s nice, yeah. I see they can write back and forth. Cool.

DC: We did this once at an institute. We passed the note around the entire table and everybody was a reacting to the person before them. It didn’t take long, but it was really interesting and kind of fun to read what everybody else wrote. All right…second helpings.

MC: All this is close reading. This is what we call close reading. Read it and then re-read it.

1:18:01 DC: Where do you stand? Take a position.

MC: Oh, yes. This is very true. You know what, I can use these for my introduction to opinion unit. We have been going back and forth about how we were going to do this. Some people wanted to do an infomercial. I said, “I don’t know.” I like this idea because it’s up to them. It’s not…but maybe we could do an infomercial with student choice.

DC: Well, ultimately giving students a choice will make it more meaningful for them in the long term.

MC: They will love it.

DC: What time of year do you teach Intro to Opinion?

MC: Now. We are late. We are supposed to have started already. This was the missing link. I think I’m going to create an activity. Then it will be more meaningful. How would you support your opinion? Where do you stand? I know it’s not meant to be that way, it just took me there. This is a great after-reading kind of thing.

DC: And I think all of these will be personalized to make them work with your group of students. And so, however you adapt it will make it valuable for you.

MC: Yes, they have to take a stand, and they have to support it. I can introduce my opinion, give my reasons. And then they can write about what they want.

DC: Right. Too many students have an opinion without any evidence. They don’t know why they like it, they just heard that someone else liked it. Now they have to come up with reasons why.

MC: We don’t do it now, but I like it.

DC: How about the RAFT essay?

MC: No. I think I’ve heard it before. Ah, for comparing texts. Wow. This is complex.

DC: This may be something more appropriate in the jr. high level.

MC: This may work once a year. But there isn’t a lot of time for this at our level.

DC: I love password. I’ve seen this done a few different ways.
MC: Is this a game show? I haven’t heard about it.

DC: Jimmy Fallon does this a lot.

MC: Yes, so they describe the word, and kids have to guess the word. Oh, I love that game. I never thought of playing in the classroom. This is nice.

DC: It’s becoming more popular lately because Jimmy Fallon does this a lot.

MC: You could use this in music.

1:23:12 DC: Yes, I think we could. And maybe even with fingerings, or playing notes, or different ways to get students to memorize patterns. This is doable.

MC: Or even with vocabulary terms. Would it work if you played a note and students had to identify the note?

DC: It's easier if they can see your fingerings, it would be harder if they could see your hands. But yeah, I could see that working.

MC: This is a great book.

DC: It is one of the best I have seen. Thank you so much for your time and your help with this project. I really do appreciate it!
APPENDIX H

PRE-EXISTING DIFFERENCES BETWEEN GROUPS AND FOURTH-GRADE MEAN ON THE MAP READING RIT SCORES
## Pre-existing Differences between Groups and 4th Grade Mean on the MAP Reading Test

<table>
<thead>
<tr>
<th>Test Type</th>
<th>Overall Reading Test</th>
<th>Literature Subtest</th>
<th>Vocabulary Subtest</th>
<th>Informational Text Subtest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pretest</td>
<td>95% CI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
<td>MD</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>t</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p</td>
</tr>
<tr>
<td></td>
<td>104</td>
<td>194.83</td>
<td>13.41</td>
<td></td>
</tr>
<tr>
<td>Overall Reading Test</td>
<td>4th Grade Total</td>
<td>4th Grade Total</td>
<td>Control</td>
<td>Group 1c</td>
</tr>
<tr>
<td></td>
<td></td>
<td>104</td>
<td>195.81</td>
<td>193.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>14.12</td>
<td>13.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>192.36</td>
<td>12.78</td>
<td>-2.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.597</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>199.36</td>
<td>5.80</td>
<td>4.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-5.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.716</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.475</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>189.80</td>
<td>12.38</td>
<td>-5.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-9.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.140</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.889</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>191.27</td>
<td>14.33</td>
<td>-1.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.686</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>199.00</td>
<td>6.59</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.866</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>190.50</td>
<td>15.74</td>
<td>-5.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-15.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.263</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>191.27</td>
<td>14.33</td>
<td>-2.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-11.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.594</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>199.00</td>
<td>6.59</td>
<td>5.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.195</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>188.70</td>
<td>12.52</td>
<td>-4.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-13.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.277</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>192.09</td>
<td>12.91</td>
<td>-2.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-11.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.631</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.529</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>200.45</td>
<td>7.45</td>
<td>5.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.115</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.236</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>190.20</td>
<td>11.77</td>
<td>-4.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-12.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.321</td>
</tr>
</tbody>
</table>

Note: M = Mean; SD = Standard deviation; MD = Mean difference between the group pretest mean and the 4th grade class mean; CI = Confidence interval; LL = Lower limit; UL = Upper limit.

a Confidence intervals calculated on the mean differences between groups; b Degrees of freedom equal 113 for the control group and group 1, and 112 for group 2. c Conventional band group. d Literacy-enriched band group.