7-12-2024

Development and initial validation of the Musical Discrimination and Styles Task: Measuring children and adolescent music aptitude and achievement

Dawn R. Mitchell White
University of South Florida

Development and Initial Validation of the *Musical Discrimination and Styles Task*:

Measuring Children and Adolescent Music Aptitude and Achievement

by

Dawn R. Mitchell White

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a concentration in Music Education
School of Music
College of The Arts
University of South Florida

Major Professor: Clint Randles, Ph.D.
C. Victor Fung, Ph.D.
David Williams, Ph.D.
Robert F. Dedrick, Ph.D.

Date of Approval:
July 12, 2024

Keywords: music aptitude, music achievement, musical discrimination, internal structure, reliability

Dedication

I dedicate this dissertation to the love of my life, Ron, our three beautiful sons, John, Jason, and Justin, Mother-in-law Diana, and Sister Ashley, and my nephew, Cody, his wife, Crystal, and their little angels—without all their love and patience, none of this would have been possible—Because of their presence in my life, every day is worth living.

I also dedicate this work to my loving parents, Jimmy and Nadine, who always believed in me; to my second dad, Dave, who gave me guidance and a loving home when I needed one; to my second mom, Joyce, who loved and supported me through thick and thin; to Uncle Joey, who believed college was meant for me even before I did; to my strong and caring role models, Aunts Sandie and Katie; to my beautiful sister, best friend, and musical soul partner, April; to my brave and beloved brother, Johnathan; and finally, to my ever-faithful and loving writing companions, Good Girl and Anakin. May God hold you all forever in the palm of his hand.

Acknowledgments

While working on my dissertation I looked at many examples of other Acknowledgements sections of dissertations belonging to other people. They were unique to each person, and I gathered that this section gave a short window into the thinking of each doctoral candidate. They varied in length and structure. Some were in first person (as is mine), and some remained in the strictness of third. Some thanked just one or two people they credited with helping them get where they were; others listed everyone they had ever known who helped them.

When I finished with my introspection, I realized that my success was not only my own. This section is essential because the people I recognized here have inspired me along my Hero's Journey. I'm older than many doctoral candidates who enter this program, so I've experienced a few more years of life's path and how it can twist when you are ready to turn in a different direction or move you forward one step while taking two steps backward. Most of my successes in life often required grit, perseverance, and stubbornness to achieve them, which made me treasure them much more (but also made me wonder about my karma 😊). I see this degree as an accomplishment I have wanted for the last twenty-five (25) years, and it's almost done.

I faced many trials as I pursued this degree (some of the most painful life lessons I've endured so far), and the people I mention in this section were integral muses along my path who inspired me to keep going when it would have been so much easier to give up. Some may not have known they were a source of strength for me, but I hope they may have known how much I respected them. If not, this acknowledgment section allows me to express how much I appreciate

them for their support, guidance, and friendship. This recognition in my acknowledgments section means that I will continue to respect them and be inspired by their examples long after I have finished this degree.

First, I want to express my sincere gratitude and respect for my major professor, Dr. Clint Randles. He has guided me through my dissertation and growth as a research musician. This program was supposed to be only three years, but there were more real-world trials than I could have anticipated when I started this degree. With all the personal twists and turns I experienced during these five years as a Ph.D. student, his insight gave me the courage, confidence, and persistence to see things through to completion. I admire his strength of character, depth of knowledge as an educator, prowess as a practicing musician, and ferocity in research. I am in awe at the level of work he can produce, and I hope that I will eventually be able to establish a creative output that is at least half of his—for I know that should set me up quite nicely to achieve tenure! [smile] Quite seriously, though, I am truly grateful for his time and guidance in my success.

I have participated in some of this journey's most enjoyable and enlightening conversations with Dr. C. Victor Fung. He taught me to be thorough, thoughtful, and curious in research. After reading his music philosophy and observing the manner and character he has revealed as he has practiced his research and music these last five years, I feel inspired to strive for his levels of balance, patience, and harmony. I owe him tremendous gratitude for my work on this project, especially since I first developed the MDAST under his tutelage. His pertinent questions and suggestions helped me bring it to fruition and refine it. Without his Measurement class, encouraging demeanor, knowledge of measurement, and insight regarding the potential of my test, this dissertation would not exist. I sincerely appreciate all his brilliant advice since I

started this degree. I am also (yes, purposely in present tense) interested in music's philosophical branch of research that combines my love of Asian culture and martial arts (I'm a 4th Dan Black Belt in Tae Kwon Do) and wrote one of my most enjoyable papers on the subject for my comprehensive exam. His book, *A Way of Music Education: Classic Chinese Wisdom*, inspired me to write the paper, Daoism and the path of the music education student. That paper is one of my favorites, and I long to refine it as soon as this dissertation is complete. I know I would never have found the courage to write that paper without Dr. Fung's influence, and I am very grateful to him.

Before I met Dr. Jennifer Bugos, I never realized that a musician could also be such an outstanding example of the scientific method. The MDAST was born the semester after I took her Music Cognition class. It helped me focus my mindset on logic and made me "ready" to create the MDAST. She set me on a path of attention to detail, protocol, and procedure. I am grateful for her time teaching me to be a more thorough researcher. I continue to be inspired by how she affects change in musical science with her careful insight, experiments, and experience. I aspire to have that same sincere dedication to my work and to effect positive changes wherever my path leads me.

My first introduction to the doctoral program at USF occurred in Dr. David Williams' office. His description of the Ph.D. program made me realize I was in the right place to learn an exceptional music education philosophy. I could tell from his smiling eyes that he loved teaching music. Later, when I had the opportunity to delve outside the Ph.D. program and take his Learner-Centered Music Education classes, I learned a teaching philosophy that has altered how I run a music classroom and will continue to run it from this point forward. I have also incorporated this teaching philosophy into a conceptual model for Twice-Exceptional students

(students who are gifted and also have a disability) and have incorporated the model into presentations at several US conferences and the ISME conference in Brisbane, Australia. I am hopeful that the conceptual model I developed will eventually help music educators teach twice-exceptional students in a way that celebrates their strengths and supports their challenges. I am so grateful to Dr. Williams for teaching me how to be a better educator, and I also hope my future students will be happier and learn more from what he taught me.

I am deeply indebted to Dr. Robert Dedrick for patiently teaching me how to write a proposal and for tutoring me in the sophisticated statistics accompanying Advanced Educational Measurement 1 and Advanced Educational Measurement 2 in research. The statistical methods were complex, yet he encouraged me to be patient and keep trying. I appreciate the extra Microsoft Teams meetings he provided to explain things to me. Without them, I do not think I would have absorbed it. When I experienced personal tragedy (the loss of my father from COVID, then the loss of my mother in a car accident almost a year later), he was the first person to extend kindness and understanding to me. Without it, I might have given up. So, beyond his intellect and outstanding expertise in educational research (which is considerable), I am grateful that he is also a kind human being.

I am incredibly grateful to have earned a cognate in “Instrumental Conducting” with two of the most outstanding conductors and people I know. First, I extend my warmest thanks to Dr. William Wiedrich. I know how fortunate I am to have studied under Dr. Wiedrich off and on over the last 28 years. It seems surreal, in a way, that I met him when he came to audition for his position at USF at the beginning of our USF careers; now, we are leaving together as he retires from USF, and I am finishing my Ph.D. in Music Education. I deeply respect him and cannot even begin to quantify everything I have learned under his instruction. He is a consummate

musician who knows exactly how to strike the right balance and tone between knowledge and intuition, patience and risk-taking, and kindness and strength. As such, any musician he taught is better for knowing him. I hope to follow his example in musical excellence, preparedness, work ethic, attention to detail, and that precious musical balance. I also extend heartfelt gratitude to Dr. Matthew McCutchen, who kept me grounded in musicking but always looking forward to future possibilities. I am so grateful for his professionalism, decency, kindness, and patience while leading me to a standard that will make me more competitive. Both outstanding conductors challenged me to be a better musician who will strive for higher expertise when I stand in front of an ensemble. They also taught me the superpower of doing my homework before stepping on the podium. When I felt overwhelmed by work or life's tragic and inevitable loss (in the middle of COVID), they encouraged me to believe in myself and gently pointed me toward success. They helped me with this dissertation by teaching me to remain grounded by finding joy in music study and interaction with an ensemble. They showed me that a quiet respite can exist when I need it with my baton and a great score. In the years to come, I'll take those noble lessons with me to share with future students, and the musical circle of life will continue.

This section would not be complete if I did not recognize the brilliant mind, gentle soul, and musical force that is Professor Ann Hawkins. I took my undergraduate music theory classes at HCC before I came to USF. Still, I had the joy of taking advanced theory classes with her during my junior and senior years of my bachelor's degree. I enjoyed those classes so much! When the time came to choose what I would do next, I chose to earn a master's degree in music theory. That gave me even more advanced music theory classes to take that would propel me forward. Simultaneously, I also enjoyed teaching undergrads in Theory 1 and 2 with Prof. Hawkins as my supervisor. I think she was the best boss I ever had! Her instructions were always

fascinating (at least they were to me!), she never expected us to do anything she wasn't willing to do herself, she treated us as equals and with kindness all the time, and we were constantly learning! I was always amazed by how much and fast she could process music. She pushed us to be like her, and all these years later, I can process music the way she taught us. It has stuck with me, and as a conductor, I find that it gives me a way to understand the music and connect with it that's "just mine." I pull the music apart in my "own" way and reassemble it with a new understanding of how it works. I can't quite explain what I mean, but I know that my most profound knowledge of music, whether tonal or any other kind of harmony, comes from the foundation I learned from Professor Hawkins. I will never forget the contributions she gave me with her patience and kind wisdom. I am grateful that I had the opportunity to develop my musical foundation from such a fantastic person!

Dr. John Robison fascinated me since I first set foot on USF soil. The first "C" I ever earned in my academic career was in his class (Medieval-Renaissance Music History). When I was younger, I had an eidetic memory for everything I saw in print but not for everything I heard in lectures. He tested us primarily on the lecture material, so I struggled to get that "C" in the first grading period. During the second grading period, I asked as many questions as I needed to understand the subject matter and took copious notes of his lectures to earn my "A," but it was clear that I had met my match in the battle of mental recall. What surprised me was that he was like a repository for music history knowledge! As long as I have known him (since 1995), he has been excellent at historical recall, musical style and ornamentation, and the correct way to play instruments from the Medieval to Classical periods. He taught me much about playing recorders from the Medieval/Renaissance in his Collegium Musicum class. I also played in that ensemble on the Dulcian (a predecessor to the bassoon). Through his tutoring, I learned well enough to

play in a recorder ensemble that got paid every year when we played for a Madrigal Dinner for the vocal department. Later, when I opened my school, the Center for Education School of the Arts and Sciences, I kept that Madrigal Dinner tradition going for 16 years with the whole student body. The students played in mini ensembles, sang, acted in short skits, had a king's table with a full royal court, and I played the lute to entertain the partygoers. Thanks to Dr. Robison, 16 years of students experienced this as a part of their childhood every year, and their families attended it as well. They learned to love the music of the Medieval and the Renaissance (which I also learned to love dearly despite my "C" grade all those years ago), and he was one of the primary inspirational reasons why all those students did. So, thank you, Dr. Robison. It couldn't have happened without you!

More than any music educator I've ever known, I want to acknowledge my first music teacher at Bay Crest Elementary School, Ms. Betsy Bouhuizen. She recognized my love for music and my endless obsession with singing and playing my child-sized guitar when I was in 1st grade. Until then, I had never received any private lessons (I could read by the time I was three, so I was teaching myself from guitar music books my uncle had), so Ms. Bouhuizen gave me private lessons for $0.50 weekly because my family could not afford to pay more. I sang and played guitar in the school chorus, the "Bay Crest Bunch," for three years. In my summer transition from 5th to 6th grade, she advised me to choose the oboe, French horn, or bassoon as my band instrument because they were excellent college scholarship instruments. I followed her advice, and all these years later, after 45 years of bassooning, I am on the precipice of completing my dissertation and earning my Ph.D. in music education. Her loving and dedicated success in the music classroom led me to follow in her footsteps and beyond. She inspired me to believe in myself, even when it was difficult to see a future beyond the poverty I grew up in. She

helped me begin my quest to be a musical performer and educator, which led me to obtain the first high school diploma ever earned in my family.

She has passed away since then, but I like to think she looks down and sees what I'm doing. She was a fantastic educator, inspiration, and leader. She made an incredible difference in the world throughout her extensive teaching career and exponentially through people like me who studied under her. Through us, she has inspired generations of music students. I am confident I would not be the woman, educator, or musician I am today without her. I hope that one day, I will have some students think of me as fondly as I do of her and may even choose to make music their career. Then again, I feel the teaching wheel of music will have come full circle.

Finally, I am thankful to have shared this journey with colleagues who allowed me to share my ideas and listen to theirs. This journey would not have been the same without them as active participants, and I cannot imagine completing it without them. I am fortunate to have entered the Ph.D. Music Education program as their peer. Thank you, Amber Alderman, Darbyleigh "Darbi" Lamrani, John "Tosh" Sargeant, and my dearest friends, Cancan Cui and Wen Zhong. I look forward to sharing the future with you, and when I choose to look back on what we have done, it will always be with great thanks and admiration.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation had two purposes: 1) to create and document the development of a new music achievement test entitled the Musical Discrimination and Styles Task (MDAST), and 2) to describe the strength of the evidence supporting the validity and reliability of this new developmentally appropriate music aptitude and achievement instrument. I created a theoretical framework based on 1) the cognitive theory of child development of Jean Piaget (1969), 2) the phase model of artistic development by David Hargreaves (Hargreaves & Galton, 1992), and 3) the theoretical models of music discrimination, audiation, and music achievement (Gordon, 1970). The MDAST's design assessed students' abilities to evaluate pitch and rhythmic discriminations and compare musical contours (all three commonly used in musical assessment), composers, and styles (new addition here but based on empirical evidence). The items were developed with the assistance of an expert panel. Following pilot testing of the created pool of items, the MDAST was reduced to 15 items organized into five subtests. Items were scored 0 (incorrect) and 1 (correct). The following research questions guided this research:

1. What is the strength of the evidence supporting the validity of the Musical Discrimination and Styles Task (MDAST)?
    a. Content validity evidence as provided by a panel of experts?
    b. Internal structure validity evidence as provided by exploratory factor analysis?
    c. Relations to other variables as provided by examining the relationships between grade level and the subtests?

2. What is the strength of the evidence supporting the reliability of the Musical Discrimination and Styles Task (MDAST)?
    a. Internal consistency reliability as provided by the Kuder-Richardson Formula 20 (Cronbach's alpha)?

Three hundred sixty-two (*n*=362) students from a community charter school in the southeastern part of the United States took the 15-item test in Qualtrics from September 13, 2022, to October 13, 2022. Confirmatory factor analysis tested the five-factor measurement model for the MDAST.

First, the researcher focused on "Descriptive Statistics" by calculating the "Item Difficulties by Total Group and by Grade Levels" [Table 4] and a "Subtest Correlation Matrix" [Table 5]. This study also established content validity through peer expert reviews who took the test and measured successful items at 80% agreement [Table 6]. Then, the researcher used confirmatory factor analysis to evaluate the five-factor model underlying the MDAST for internal structure validity. The researcher then assessed reliability using the Kuder-Richardson Formula 20 on both the total test and the subtests [Table 7]. The results for the overall test revealed for all participants was $\alpha = .681$. For the subtests, their results were as follows: 1) Pitch $\alpha = .449$, 2) Rhythm $\alpha = .398$, 3) Contours $\alpha = .118$, 4) Composers $\alpha = .346$, and 5) Styles $\alpha = .056$. There were implications for future research, as well as those for current practice.

Chapter 1: Introduction

Since music is an accepted form of education, and accountability is necessary for all facets of education, it is readily apparent that there will always be a need for forms of music student evaluation. One of the most important actions I took in this study was to differentiate and choose between aptitude and achievement and determine how the selected assessment type best represented the test I wanted to create. Aptitude testing evaluates a student's inherent musical ability before training. Those tests represent music learning potential (Conway, 2020). Conversely, achievement testing discovers what a student has learned up to that particular moment in time. It is the kind of testing most often used to assess levels of absorption and attainment after coursework (Conway, 2020). There are also standardized musical assessments that measure student achievement on a larger scale.

The principal investigator (me) created the MDAST. I intended this study to contribute to the musical body of knowledge with this aptitude and achievement instrument. I also aimed for the MDAST to parallel students' musical learning outcomes throughout their elementary academic careers. I can see how new pathways could open up in this field of research.

1.1 Two Purposes to Serve as a Guide

I established two purposes for this study to guide all work throughout this dissertation. Every action I took had to be placeable "under" one of these two purposes to remain a part of the study. This standard was my way of assuring that there would be consistency throughout the creation of the assessment, the analysis, and the writing of this document. For the **first** purpose,

1) I created and documented the development of a new music aptitude and achievement test entitled the *Musical Discrimination and Styles Task* (MDAST). **2)** The **second** purpose described the strength of the evidence supporting the validity and reliability of this new developmentally appropriate music aptitude and achievement instrument (The students took the MDAST, and I described the results in "Chapter 4"). Documentation and experimentation are needed to remain active in all planning, writing, analysis, and interpretation throughout this study.

The **first** purpose of this dissertation instrument occurred when I *began to document* the MDAST, and it was fulfilled a little more whenever I worked on this dissertation. The **second** purpose was set into motion as soon as I started to *explain* the strength of the evidence supporting the validity and reliability of the MDAST. Before I could explain the final validity and reliability findings, I had to search for evidence. Once I had my evidence (data), the actions I took and the records I kept led to the most important part of the second purpose. The essential finale was the necessity for me to interpret the results and present my findings in this document.

1.2 Background, Context, and Theoretical Framework

For music researchers and educators, understanding how students grow and demonstrate their musical aptitude and achievement is critical (Gordon, 1965). To achieve this level of understanding, the researcher must use good theory, and conversely, psychometrically sound measurement methods must be used to test and construct theory. Theory and measurement are inseparably conjoined. In the development of sound instruments, sound theory is required, and in the development of sound theory, sound instruments are necessary. Therefore, I present three essential theories, those of Jean Piaget (1957), Edwin Gordon (1970), and David Hargreaves (1966), as the cornerstones of the MDAST. See [Figure 1] on page three.

Elementary music students undergo exponential development and understanding stages, influencing their musical ability. As a result, Jean Piaget's theory of four major, qualitatively distinct stages of cognitive development (Piaget, 1972) is one of the main influences on the Musical Discrimination and Styles Task (MDAST), a test of musical aptitude and achievement.

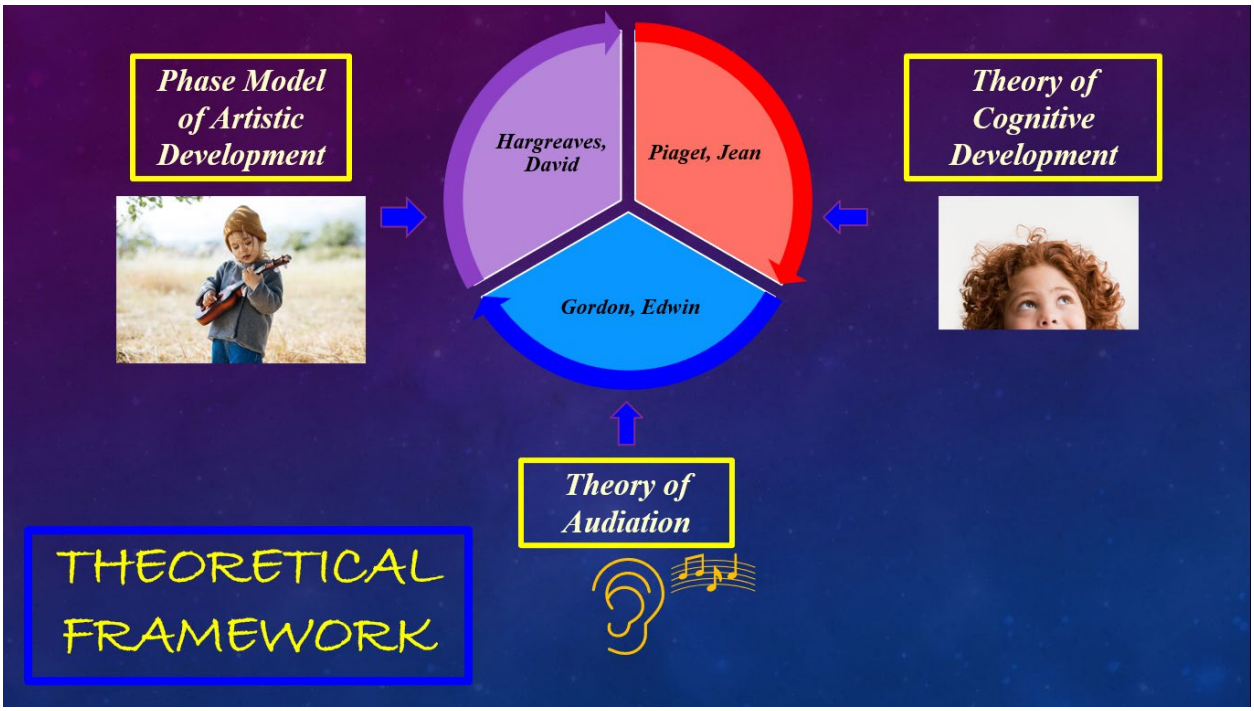

Figure 1. *Theoretical Framework Diagram*

He describes these four stages in his book, *The Psychology of the Child* (1972). The preconceptual phase (ages 2–4 years) and the intuitive phase (ages 4–7 years) are two subphases of the second pre-operational stage (ages 2–7 years). According to Piaget's stages, some younger children within these phases (especially in the lower range) might not be ready for the experiences within this study. However, Piaget described that children enter the concrete operational stage at about the age of seven. This stage directly applies to the MDAST study. In this stage, children become more logical thinkers and are less egocentric. Significantly, their

logic is limited to concrete ways because the children at this stage are not developmentally ready for abstract ideas. But they are ready *for concrete use of logic in tangible ways*, such as items that can be **touched, seen, or experienced directly with the other senses**. At the age of 11 years, *abstract systematic organizational* thought emerges. Hence, children can have extraordinary musical growth during the concrete operational stage, leading to age 11.

Hargreaves and Galton (1992) based their Phase model of artistic development on Piaget's Stages of cognitive development, and I adopted it as the study's scaffolding. Within it, children can identify global features in music, such as *pitch and contour*, during their Figural Phase (ages 2–5 years) under Melodic Perception. The Schematic Phase (ages 5–8 years) teaches children about *melodic property conservation*. Finally, children undergo *analytical knowledge of intervals* and *key stability* during the Rule Systems Phase (ages 8–15 years).

The Phase model of artistic development (Hargreaves & Galton, 1992) and the theoretical Model of cognitive development (Piaget, 1972) are two of the three developmental partners forming the conceptual basis for this measure and research. As a result, the model used in this study focused on a child's ability to *distinguish between notes* (Cooper, 1994), determine *"same or different"* between *rhythms*, *assess contours* (Fancourt et al., 2013), and *identify related composers* (Ilari & Polka, 2006), and *styles* (Addessi et al., 1995).

1.3 MDAST - Determination Between Aptitude and Achievement

One of the most critical decisions throughout this study was determining the MDAST's status as an aptitude or an achievement instrument. Therefore, I researched to sort out which kind of test I had actually created. I had several instances over the last few years where this status was called into question (even as I was preparing for my dissertation defense) when I needed to

reevaluate my research and come to an informed conclusion. This section is an answer to this critical question: aptitude or achievement?

Gordon (1999) asserted that musical aptitude results from innate potential and early experiences in a child's environment. They combine in unknown quantities to create music aptitude, and the amount of each factor needed is unidentified. Furthermore, Gordon (1999) explained that despite the level of musical aptitude a child is born with, they require both formal and informal musical experiences to maintain and build upon that level of talent and potential. Without these experiences, musical aptitude will never reach full realization in musical achievement. Gordon (1999) stated, however, that a child's early experiences, neither formal nor informal, can push their musical aptitude to a higher level than they had when they were born. According to Gordon (1999) and Seashore (1915), developmental music aptitude stabilizes at age nine and stays the same for a person's lifetime. As regards the MDAST, the tonal and rhythmic subtests are almost always used in aptitude instruments, but the ability to assess contours, identify related composers, and differentiate between musical styles is not. I have not found any aptitude test that includes those subtests.

Achievement testing verifies a participant's level of success within areas in a domain after learning, training, or practice. De Manzano and Ullén (2021) conducted a study in which they needed to investigate and identify personal and cultural factors that foster creative musical achievement. Their study recognized that music achievement depended on general and specific abilities (aptitude) that were affected by the music domain-specific, personality traits, and childhood influences. Childhood environment is explicitly mentioned in both studies of aptitude and achievement for continued development. Therefore, Achievement is affected and determined

by the participant's success after learning, training, or practice, but is also affected by the childhood environment, their aptitude, the music domain, and their personality traits.

Therefore, I assert that the Musical Discrimination and Styles Task (MDAST) is both an aptitude and an achievement test because it contains elements of both an aptitude and an achievement test. It contains the pitch discrimination and rhythm discrimination subtests that can be found in almost every aptitude test. It also contains three subtests of items that can be found in the childhood environment: musical contours, music of composers, and musical styles. While these last three do not require the test taker to know the names of the composers or the musical styles, they will have been exposed to these kinds of music in their environment, and they will have to make a determination based on that level of music exposure and culture. As they grow older, they will be exposed to more kinds of music, expanding their knowledge, both in their environment and in their music classes.

1.4 Development of a New Music Achievement Test Based on Older Ones

Edwin E. Gordon's (1979) aptitude tests and his achievement test were influential in creating the MDAST test since they centered explicitly on the principle of audiation (Gordon, 1989). According to him, audiation was the cornerstone of musical aptitude, which in turn was the foundation of musical achievement. He described audiation as hearing and understanding music without physical sounds present (Gordon, 1989). He described seven types and six stages of audiation, some of which applied to the MDAST.

Only *Type 1* (Gordon, 1989) of the seven forms of audiation was relevant to the MDAST. It happens when people *listen* to music that is either familiar or new to them. During the listening activity, the individual connected and identified both tonal and rhythmic patterns that

they heard in the music but are no longer physically present to provide syntactical meaning to the music they heard through the audiation process. This audiation type fits its definition because the MDAST contained items that involved a response to listening examples. I built the MDAST to align with the first of these six stages of audiation; its expression was in its architecture.

1.5 Problem Statement

While researching musical achievement tests for this study, I found fewer than I had hoped. Most of the tests I found were geared only toward music aptitude. Of the musical achievement tests I found, I was disappointed to find that they were only testing for pitch and rhythm competency. For instance, the Iowa Tests of Music Literacy (Gordon, 1971) stands out as one of the best-known music achievement tests. Despite being divided into six subtests, they still fall under two headings: Tonal Concepts and Rhythm Concepts. Gordon did not consider any other music achievement areas. From the music achievement hunt and the results I discovered (as above), I created a problem statement to explain why I chose to create the MDAST the way I did.

I decided to craft this problem statement using a *four-elemental approach* (Applied Doctoral Center, 2024) to explain the research conducted in this dissertation. 1) The ***first*** element in this approach was *to specify and describe the problem* studied. Therefore, I began by elaborating on the first aspect of the problem.

I created the MDAST because my research demonstrated that, historically, music aptitude and achievement assessment fields were mostly centered and dependent on the audiation of pitches and rhythms. As time progressed, these tests still focused primarily on those two concepts: the identification of pitches and rhythms. Here's where 2) the ***second*** element in the

problem statement occurred: the revelation of the *evidence of the problem's existence*. Once again, the prominently known achievement test by Edwin E. Gordon (already mentioned at the beginning of this section), the *Iowa Tests of Music Literacy (ITML)*, serves here as an example of the second element. This test was a normed reference test for music achievement. However, it still *only contained two subtests:* Tonal Concepts and Rhythm Concepts. 3) The ***third*** element of constructing a thoughtful and thorough problem statement was to *explain the consequences of NOT solving the problem*. If researchers continued on a path that only addressed two subtests in a test population with so many opportunities to choose from, then it would create an atmosphere of stagnancy. It is essential to move forward and find new approaches to teach, analyze, develop, re-invent, and, sometimes, re-evaluate our relationship with music to keep moving forward. The cost of NOT searching for new approaches to evaluate student learning is critical. As we know, in academia, there are many different learning styles that our students can learn and engage with. Suppose we continue to accept that pitch and rhythm are the "accepted" ways to assess a student's musical aptitude and achievement. In that case, we neglect other ways students can manifest a relationship with music. In addition to determining a student's music aptitude, the MDAST is an attempt to determine when students successfully demonstrate their ability to recognize music in a way they have learned to do since early childhood.

This brings the 4) ***fourth*** element to this music problem: *to identify what is not known about the problem that should be known.* In the MDAST, I endeavored to assess student aptitude levels in pitch discrimination (tonal) and rhythmic discrimination (rhythm) subtests. However, I also sought student achievement levels in musical contour, composer, and musical style comparisons. These three new subtests allowed students to demonstrate their ability to make comparisons based on previous knowledge gained through their music learning. As a result, the

finished model I used to create the MDAST in this study focused on a child's ability to **(Subtest 1)** *distinguish between notes* (Cooper, 1994), **(Subtest 2)** determine *"same or different"* between *rhythms*, **(Subtest 3)** *assess contours* (Fancourt et al., 2013), **(Subtest 4)** *identify related composers* (Ilari & Polka, 2006), and **(Subtest 5)** *differentiate between musical styles* (Addessi et al., 1995).

Additionally, as students progress through their music education, they are exposed to greater information levels within the five MDAST subtests. Therefore, the expectation should be that students' achievement levels on the MDAST grow (go up) as they progress through the grade level system.

1.6 Purpose of the Study

A need exists for a valid and reliable measure of musical discrimination and musical style aptitude and achievement that is uniquely informed by robust psychological theories of childhood development, like Jean Piaget's Model of Cognitive Development and David Hargreaves's Phase Model of Artistic Development. Such a contribution could significantly advance understanding of how children and adolescent students may eventually develop into competent musicians.

This paragraph describes a simple reminder of the overall purposes I have determined are the aims this study. For the **first** purpose, **1)** I created and documented the development of a new music aptitude and achievement test entitled the "Musical Discrimination and Styles Task (MDAST)." **2)** The **second** purpose described the strength of the evidence supporting the validity and reliability of this new developmentally appropriate music aptitude and achievement instrument. (The students took the MDAST, and I described the results in "Chapter 4.") Of

course, these guiding purposes needed to remain active in all planning, writing, analysis, and interpretation throughout this study.

My purpose typology would be to "test new ideas" (Newman et al., 2003). The survey design assessed the ability to determine pitch and rhythm discrimination, contours, composers, and styles in a "same or different" response format.

1.7 Research Questions

1. What is the strength of the evidence supporting the validity of the Musical Discrimination and Styles Task (MDAST)?
    a. Content validity evidence as provided by a panel of experts?
    b. Internal structure validity evidence as provided by exploratory factor analysis?
    c. Relations to other variables as provided by examining the relationships between grade level and the subtests?
2. What is the strength of the evidence supporting the reliability of the Musical Discrimination and Styles Task (MDAST)?
    a. Internal consistency reliability as provided by the Kuder-Richardson Formula 20 (Cronbach's alpha)?

1.8 Significance of the Study

This study could have a sustainable impact on music education. It could be used in an educational format as a possible standard test of both music aptitude and musical achievement in music classrooms. For example, when comparing the amount of time a music teacher would have for their students to take the *Iowa Test of Music Literacy* (testing on more than one session) to

the *MDAST Abbreviated* (one session of 20 minutes), the educator might feel more comfortable fitting the MDAST into their schedule. Similarly, students may feel less intimidated about taking a shorter assessment that only takes 20 minutes.

By tracking a student's aptitude/achievement scores from year to year, a music teacher can evaluate whether a student is making sustained progress as they grow from grade 1 to grade 8. However, one test score would not be enough to determine a student's aptitude/achievement over time. At least two to three assessments would be necessary to corroborate the student's growth. In situations like these, z-scores are used to differentiate between tests. The z-score would be the determining factor of the student's musical aptitude and achievement. It would also be possible to use z-scores on individual subtests, which could separate the aptitude from the achievement, to help parents and teachers determine whether a student is progressing from year to year.

If the MDAST is valid and reliable, it would be a viable measure of research capable of continuing research possibilities. Good research begets opportunities for more good research.

1.9 Gaps in the Literature

In his most extensive and well-rounded instrument, the *Musical Aptitude Profile* (MAP), Gordon (1995) tested children (Grades 4 – 12) in six subtests: ***tonal*** imagery-melody, **tonal** imagery-harmony, **rhythm** imagery-tempo, **rhythm** imagery-meter, **musical sensitivity**-phrasing, and **musical sensitivity**-balance. He created the *Primary Measures of Music Audiation* (PMMA) to assess the audiation abilities in children (Grades K – 3) in two subtests, **tonal** and **rhythm**. Gordon developed the *Intermediate Measures of Music Audiation* (IMMA) to evaluate the audiation of children (Grades 1 – 6), but once again, he only assesses two subtests, tonal and

rhythm. In his **only achievement test**, the Iowa Tests of Music Literacy, he created six subtests divided under two domains, **Tonal Concepts and Rhythm Concepts**, for students in grades 4 – 12.

However, all MDAST versions address the literature gap by testing children in five subtests. They address the *tonal and rhythmic* aspects of audiation but also represent the child's ability for *critical reasoning*, given the opportunity for *comparison and audiation of musical excerpts*. The five updated subtests are pitch discrimination, rhythm discrimination, comparison of musical contours, comparison of composers, and musical styles.

1.9.1 Musical Composer Comparison

I wanted to determine if students could discern between composers upon listening to two musical examples. Unfortunately, I could not find another aptitude or achievement test that assesses this ability, so I knew I needed to have previous instances where students/children could demonstrate this ability, even if it were outside of a formal testing situation. Hence, I researched to establish credibility for a musical composer comparison subtest.

According to Marshall and Hargreaves (2007, p. 33), "Gardner (1973, p. 326) operationalized style sensitivity in music as the ability to recognize whether two contrasting musical pairs came from the same or a different piece of music, that is, as 'the ability to group together works produced by one artist.'" While Gardner initially discussed musical style, he supported another important musical ability with this comment. If a child can group works according to whether or not a single artist (or composer) produces them, they can differentiate when works are not by the same artist (or composer). Children can compare and contrast pieces of music against one another and make a value judgment regarding their status. Therefore, the

child can answer the following question: Are the pieces of music composed by the same person, or are they written by a different person?

After considering the studies by Ilari and Polka (2006), Marshall and Hargreaves (2007, p. 33), and "Gardner (1973, p. 326), I decided to create some parameters for the creation of the musical composer comparison subtest examples. *First*, when I wanted to create an example for a composer to be recognized as themselves, I would collect both music excerpts from the same piece of music and the same movement. I would collect a familiar melody for the first excerpt. For the second excerpt, I would try to find a place in the music where the composer was developing a theme heard in the first excerpt.

Second, if I wanted the composers to be recognized as different, I would not collect excerpts from the same period or genre together (e.g., Mozart and Haydn). Rather, I would select excerpts that exemplified those differences to make it easier for the ear to recognize the difference in composer (e.g., Bach's Goldberg Variations and Beethoven's Fifth Symphony, First Movement). I made a special point of using a few world music excerpts and a few contemporary ones, as well. Some of the students in today's classrooms are from different cultures around the world, and some are engaged in listening to world music online. Many students are on social media and listen to music from a plethora of sources, artists, genres, styles, and eras. Many are also young children who might be unable to relate to all those resources. I tried to strike a balance between the target group (grades 1 – 8).

1.9.2 Musical Styles Comparison

Gardner (1973) found that first graders had a positive style response and that the mean scores at the upper age levels were high in his musical style study. He partially credited his task's

performance to the gifted target population he had selected for the report. He speculated that a group of less intelligent or motivated children might have produced different results (Gardner, 1973). “An important consideration in perceiving styles is whether two works feel differently or have distinct Gestalten,” he suggested (Gardner, 1973, p. 75).

However, Marshall and Hargreaves (2007) suggest that many preschool children seem capable of perceiving differences in musical styles. They credit their study’s success with Gardner’s success in his pursuit of sensitivity to styles within music and the arts. Marshall and Hargreaves describe his study as “pioneering research” (Marshall & Hargreaves, 2007, p. 33). They provided a synopsis of his study, conducted with children’s interpretation and understanding of musical style.

The children in this study ranged from age six to eighteen. First, he selected Western classical musical examples from 1680 to 1960. Then, Gardner had the participants play pairs of short musical excerpts that exemplified four musical periods: Baroque, Classical, Romantic, and Modern. (A note here is that he equated musical style with a musical era. Thus, this illustrates the crossover between style and era I wanted to avoid in my study. They were too highly correlated.) Gardner’s experiment included a systematic comparison of periodic styles against one another. For instance, he compared Baroque with Classical, Romantic, and Modern, then Classical with Baroque, Romantic, and Modern (etcetera). After listening to a set of excerpt pairs, the participants had to choose whether they believed the two excerpts were from the same or different pieces of music. The results revealed that even the youngest participants in the study (age six) demonstrated high sensitivity to musical style.

In a study of musical style preferences and aural discrimination skills, May (1985) suggested a positive relationship between primary school children’s music preferences and their

abilities to make same/different decisions about musical stimuli pairs (Van Zee, 1976). Therefore, the May and Van Zee studies were positive continuations of the literature regarding children's ability to make musical style-based decisions.

1.10 Operational Definition of Terms

This section is a list of the definitions critical to understanding the dissertation.

1.10.1 Critical Terminology

1. *Aptitude* – " 'potential' or 'potentiality'—a latent, present, inferred quality or power that makes possible the development, given specified conditions, of some further quality or power, positive or negative…means aptness, inclination, tendency, propensity, predisposition, fitness, or suitability for performance in some future situation" (Snow, 1991, p. 250). "An aptitude is a combination of characteristics indicative of an individual's capacity to acquire (with training) some specific knowledge or skills such as the ability to speak a language, to become a musician, or to do mechanical work….The terms, special ability or talent, are used synonymously with the term aptitude. Aptitudes are natural talents, special abilities for doing or learning to do certain kinds of things easily and quickly. Musical talent and artistic talent are examples of such aptitudes" (Kaur et al., 2016, pp. 38–39)
2. *Aptitude test* – "An illustration of a special aptitude test would be a test that measures mechanical aptitude or electrical aptitude alone. The musical aptitude test is a special aptitude test. At the other end, multiple aptitude tests combine a set of separate tests together such that the individual tests measure relatively different independent abilities" (Kaur et al., 2016, pp. 38–39)

3. *Achievement* – acquisition, learning, or knowledge representation, sometimes depending on theoretical biases (Algarabel & Dasi, 2001, p. 45)
4. *Achievement test* – diagnosis at an individual or institutional level, in addition to accountability is a goal of achievement testing (Algarabel & Dasi, 2001, p. 44).
5. *Schema* - how children interpret and organize information through frameworks or mental structures (Widmayer, 2004).

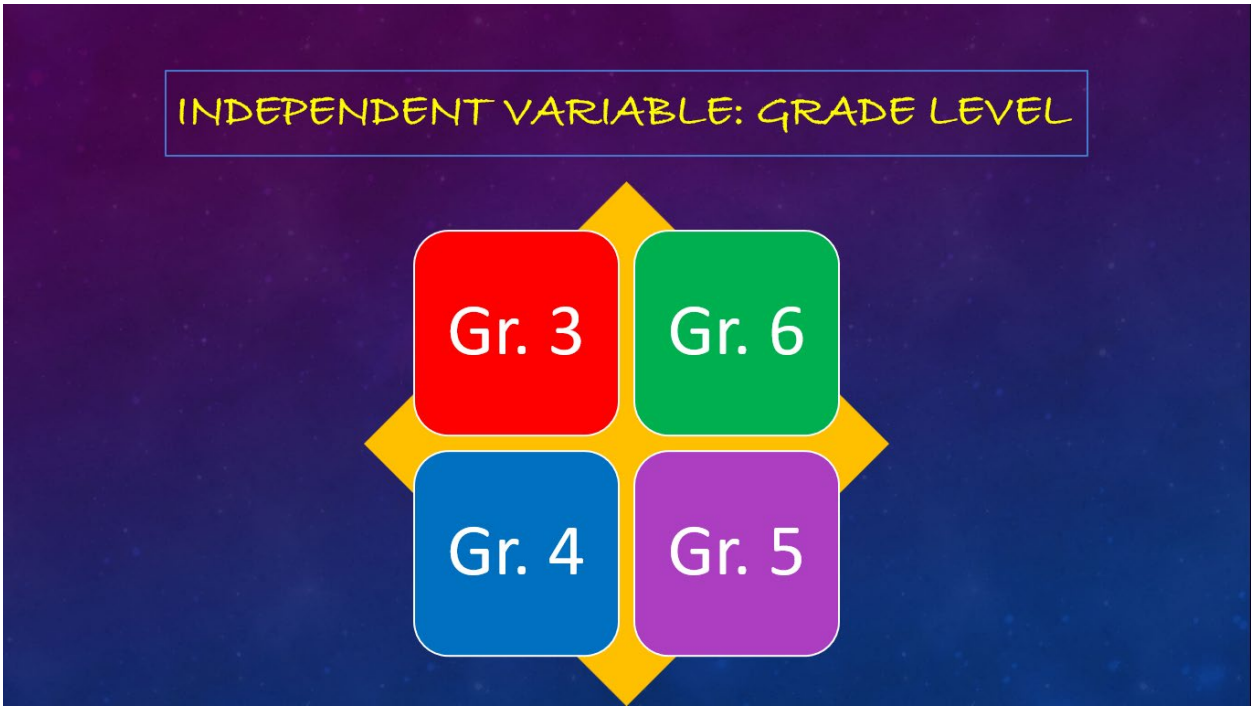

Figure 2. *Independent Variable: Grade Level*

1.10.2 Independent Variable

The independent variable was "Grade Level" when I established grade-level comparisons in the test. It was initially subdivided into the eight categories of Grades 1 – 8 to cover each of the grade levels available in the sample. However, after analysis, the Grade Levels changed to 3 – 6.

An explanation occurs later in this dissertation. The Grade Level variable is part of the "relations to other variables" validation.

1.10.3 Dependent Variable

The dependent variables are listed below. These five categories should ***not*** be strongly correlated to one another to maintain their validity. They are also a part of the "relations to other variables" validation.

1. *Pitch discrimination* – This variable refers to the composite score of all items on the 'pitch discrimination' section of the *Musical Discrimination and Styles Task*. It consists of three items. The test asks students to determine whether *two pitch examples* are the same, different, or "I don't know." Each example features two simple, pure tones, created on NCH Tone Generator and made into an example on WavePad Sound Editor. The participant hears the tones devoid of rhythm or any other musical device so that they may decide based solely on the pitch.
2. *Rhythmic discrimination* – This variable refers to the composite score of all items on the 'rhythmic discrimination' section of the *Musical Discrimination and Styles Task*. It consists of three items. The test asks students to determine whether *two rhythm examples* are the same, different, or "I don't know." Each example features two composed items, which utilized pure tones from the NCH Tone Generator and combined on WavePad Sound Editor. While listening to these examples, the participant hears both melody and rhythm. However, the rhythm is the only part that may change. The melody in each example always remains the same.

3. *Melodic contours* – This variable refers to the composite score of all items on the 'melodic contours' section of the MDAST. It consists of three items. The test asks students to determine whether *two melodic contours* are the same, different, or "I don't know." Each example features two composed items, which utilized pure tones from the NCH Tone Generator and combined on WavePad Sound Editor. Both melody and rhythm function in these examples. However, the melodic contour is the only aspect of the example that changes. The rhythm in each example always remains the same.
4. *Musical composers* - This variable refers to the composite score of all items on the 'musical composers' section of the MDAST. It consists of three items. The test asks students to determine whether two musical excerpts are from the same *composer*, different, or "I don't know." Each example features two musical excerpts for the participant to hear. The participant must compare and contrast the excerpts to arrive at an answer they believe to be correct.
5. *Musical styles* - This variable refers to the composite score of all items on the 'musical styles' section of the MDAST. It consists of three items. The test asks students to determine whether two musical excerpts are in the same *style*, different, or "I don't know." Each example features two musical excerpts for the participant to hear. The participant must compare and contrast the excerpts to arrive at an answer they believe to be correct.

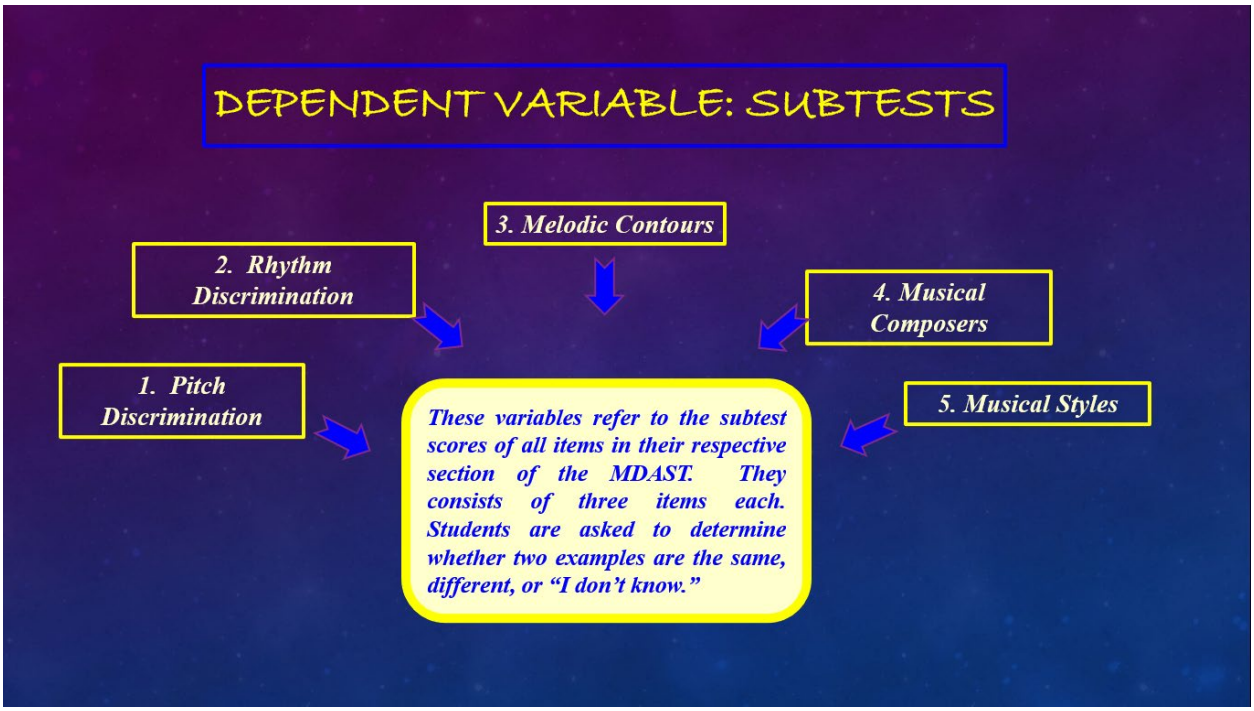

Figure 3. *Dependent Variable: Subtests*

1.11 Delimitations

This study was intended for children in grades 1 - 8 (ages 6 – 14), but ended up encompassing children in grades 3 – 6 (ages 7 – 12). It also included children who gave their online Assent and whose parents gave their *paper permission form consent* from that group.

1.12 Limitations

There were several limitations to this study. **First**, the study was limited to the sample it received from a Community Charter School in the Southeast United States. **Second**, there was no guarantee that teachers would not help their students with the test. **Third**, I chose to give the students three options to choose from when answering a question on the test (same, different, or "I don't know"). If they chose the answer, "I don't know," it was marked wrong (it was meant to

keep the student from guessing on the test). If I had chosen to simply omit those questions ("I don't know") from the student's scores, the reliability and validity may have turned out very differently. **Finally**, even though we had 362 participants enrolled in the study, it was still too small to generalize the results.

1.13 Summary of the Chapter

Chapter 1, the "Introduction," provided an overview of the essential concepts on which this dissertation relied. Next, the chapter presented the two purposes to be infused throughout this dissertation (as listed at the beginning of the chapter).

> For the **first** purpose, **1)** I created and documented the development of a new music aptitude and achievement test entitled the *Musical Discrimination and Styles Task* (MDAST). **2)** The **second** purpose described the strength of the evidence supporting the validity and reliability of this new developmentally appropriate music aptitude and achievement instrument.

From there, the chapter discussed the "Background, Context, and Theoretical Framework," which gave the reader(s) an understanding of the contributing factors that led to the creation of the MDAST. It delineated the formative theories that were pivotal in manufacturing the assessment. In "Development of a new music achievement test based on older ones," the section discusses the influences of older achievement tests on creating the MDAST. The "Problem Statement," especially, was pivotal in describing why the MDAST was needed, and it

serves as a precursor for the “Significance” and “Gaps” sections below. This chapter's “Purpose” section was closely aligned with the “aims” of the dissertation as a whole, as seen in this paragraph. The “Research Questions” were presented in this chapter to allow both the researcher and reader(s) to understand the logic behind the research and the order of conducting the research analysis to reach the aims of this study. The “Significance of the Study” and the “Gaps in the Literature” explained why this study was worthy of being conducted and where it fits in compared to the rest of the existing literature. The rest of Chapter 1 focused on acquainting the reader(s) with the “Critical Terminology” used in this study: *“*independent variables” (grade levels), “dependent variables” (subtests), and the “Delimitations” and “Limitations” of the study.

Moving forward into Chapter 2: Review of the Literature, the research focused on discovering tests that came before creating the “*Musical Discrimination and Styles Task*.” I sought music achievement assessments that stepped outside the “tonal/rhythm paradigm” and established something new in music achievement testing. Reflections on the research results are in the next chapter.

Chapter 2: Review of the Literature

This literature review was a necessary overview of the background knowledge that supports my dissertation. It presents descriptions and explanations in a logical sequence of related topics that prepared me (as the principal investigator) to write the document. When I created my measure, I had to decide the nature of its focus. I initially aimed to develop an aptitude test, but the last three subtests of my test led to a re-categorization as an aptitude and achievement test.

Throughout this dissertation, I kept reminding myself of the purposes I placed in front of me as a guide to why I conducted this research, and why I am writing this dissertation. I partially filled the **first** purpose because I documented the development of the MDAST, but the study and dissertation were not finished. As I was working on them, I would become aware of developmental adjustments that needed to be made or new measurement approaches to try that would provide me with better results. The **second** purpose described the strength of the evidence supporting the validity and reliability of this new developmentally appropriate music aptitude and achievement instrument. This purpose has been a constantly evolving job from the moment I started my pilot study, and I do not anticipate it to end. Since I will be working on the MDAST in future studies, I believe both of these purposes will become long-time friends.

I first discussed Aptitude vs. Achievement in this literature review. Next, since my measure, the MDAST, was recognized as an aptitude/achievement measure, I described the achievement measures of three well-known and respected music researchers: James Aliferis, Richard Colwell, and Edwin Gordon. After that, I discussed two common subtests in music

education achievement tests: *Pitch Discrimination and Rhythm Discrimination.* Following that, I offered a rationale for three possible subtests in Music Achievement: *Musical Contour Identification, Musical Composer Comparison, and Musical Style Comparison.* Finally, I summarized and shared my commentary on the literature.

2.1 Aptitude vs. Achievement

Aptitude testing determines a participant's inherent abilities in a specific domain. Regarding musical aptitude tests, Edwin Gordon created and evaluated more tests than most other music researchers. Gordon (1999) asserted that musical aptitude results from innate potential and early experiences in a child's environment. They combine in unknown quantities to create music aptitude, and the amount of each factor needed is also unidentified. No evidence exists to support that music aptitude is an inherited trait. Therefore, the level of a child's music aptitude is unpredictable based on inheritance. Furthermore, Gordon explained that despite the level of music aptitude a child is born with, they require formal and informal musical experiences to maintain and build upon those levels of talent and potential. Without these experiences, the music aptitude will never reach full realization in musical achievement.

Gordon stated, however, that a child's early experiences, neither formal nor informal, can push their music aptitude to a higher level than they had when they were born. No one, under any circumstances, can raise a child's innate musical aptitude level. Whether they have a high or low music aptitude, it will disappear without nourishing musical experiences to help maintain it.

According to Gordon (1999) and Seashore (1915), developmental music aptitude stabilizes at age nine and stays the same for a person's lifetime. This theory does not mean they cannot be taught music after age nine. It means that the commensurate level a person can expect

to reach in music achievement can only reach the level of their stabilized aptitude (potential to achieve). Most people do not develop their full musical aptitude to its highest level by the age of nine. Therefore, they do not perform musically at the highest level their aptitude allows.

Achievement testing verifies a participant's level of success within areas in a domain after learning, training, or practice. De Manzano and Ullén (2021) conducted a study in which they needed to investigate and identify personal and cultural factors that foster creative musical achievement. First, they identified a broadly inclusive set of variables relevant to attaining music achievement. Then, they assessed how the variables mutually anticipated the realization of a state of being (non-musicians vs. amateur musicians vs. professional musicians) and the amount of (number) achievements received among professional musicians. The variables within the set under consideration included some "general and specific abilities (general ability, auditory ability, absolute pitch), broad and specific personality traits ([extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience, also called the big-five personality traits and] psychosis proneness), childhood environment (number of musicians and recordings in the home, going to concerts), and total music practice, along with several control variables (age, sex, music genre, music occupation)" (de Manzano & Ullén, 2021, p. 2). The study results showed that music domain-specific abilities, personality traits, and childhood influences better distinguish non-musicians, amateur musicians, and professional musicians. Furthermore, those same significant predictors were related to the number of creative achievements among professional musicians.

The de Manzano and Ullén study on music achievement is in harmony with Gordon's studies on music aptitude. While it does not confirm the stabilization of music aptitude at age nine, it does recognize that *music achievement depends on general and specific abilities*

(aptitude) that are affected by the music domain-specific, personality traits, and childhood influences. Childhood environment is explicitly mentioned in both studies of aptitude and achievement for continued development.

2.2 Intro to Three Respected Music Achievement Measures: Aliferis, Colwell, and Gordon

This section of the Literature Review focuses on the respected history behind the "music achievement test" and the researchers/creators of these assessments. These researchers laid the foundation for the music achievement test with their tests, some of which are still used today.

2.2.1 James Aliferis: The Aliferis Music Achievement Test (MAT) (1957)

James Aliferis created his "Music Achievement Test" for post-secondary music students upon their first entry to collegiate music studies. This standardized test was first released in 1954, followed by the distribution of a second level in 1962 (Aliferis, 1957; Shuter-Dyson & Gabriel, 1981). Aliferis (1957) described his assessment as the "ability to hear with the inner ear what is seen in notation and to visualize the notation of music that is heard" (p.6), which can also be described as **auditory-visual discrimination**. He split the measure into two levels and used the piano as the only instrument for each musical example. The first test level consisted of the following subtests: melody, harmony, and rhythm. The second test level consisted of the following subtests: harmonic patterns, melodic idioms, and rhythmic idioms. Intervallic items represented the melody and harmony; all rhythmic examples were equal to one beat. Each idiom consisted of a four-note pattern (Aliferis, 1957; Shuter-Dyson & Gabriel, 1981). Aliferis (1957) stated that this test was comparable with abilities to dictate music and sight-sing.

Erlings (1977) used the *Aliferis Music Achievement Test* (MAT), which required the participants to complete two sight-reading piano subtests. After the participants completed all the tests, Erlings scored the MAT. One of the conclusions reached by Erlings was that the MAT correlated significantly with sight-reading achievement.

2.2.2 Richard J. Colwell: Music Achievement Tests (MAT) (1970)

Richard Colwell published his Musical Achievement Tests (MAT) in 1969 and 1970. He designed the tests specifically for students in grades 3 to 12 to measure their aural musicianship skills. These tests aimed to determine which students would be compatible with and who would benefit most from formal instrument instruction. It also attempted to improve the music curriculum of the time (Shuter-Dyson & Gabriel, 1981). Kornicke (1992) states that these tests were one of a few published tests that measured "aural imaging" (p. 154), which is essential for ear training. The importance of using ear training as an integral part of musicianship, to accompany instrumental instruction, and to improve the music curriculum was a pivotal idea of the 1970's. However, Colwell wasn't the only theorist to go down this "aural" path. In the next section, Colwell's contemporary, Edwin Gordon, also centered his musical ideology on "audiation," a similar word with a broader meaning.

The MAT contains four tests within its design, with options to either administer as a group or as standalone tests. (Shuter-Dyson & Gabriel, 1981). Researchers have often distributed Test 2 with tests from other sources (MacKnight, 1975; Palmer, 1976; Miller, 1988; Levy, 2001). The test 2 design contains three subtests: *major-minor discrimination*, *feeling for tonal centre*, and *auditory-visual discrimination* (Shuter-Dyson & Gabriel, 1981). For the complete

Colwell MAT, there is high reliability with the complete score of all tests (Shuter-Dyson & Gabriel, 1981).

2.2.3 Edwin E. Gordon: Iowa Tests of Music Literacy (ITML) (1970)

"The Iowa Tests of Music Literacy are designed to sequentially assess basic music achievement in tonal and rhythm audiation and notational audiation" (Gordon, 1970, 1991, p. 1). These were Gordon's only standardized tests of music achievement. Uniquely, he defined notational audiation as the "ability to comprehend, for example, the tonality, meter, and functions of the tonal patterns and rhythm patterns being read and written in music notation" (Gordon, 1970, 1991, p. 1). It played a pivotal role in developing these stabilized and standardized tests of musical achievement for Grades 4 – 12 by introducing a written element into one of his tests.

The Iowa Tests of Music Literacy (ITML) (Gordon, 1970, 1991) contain six levels designed to assess parallel concepts across them. Each level includes a measure of six sub-tests, which fall under two categories: Tonal Concepts and Rhythm Concepts. Within each category are three test divisions: Audiation/Listening, Audiation/Reading, and Audiation/Writing. Levels 1 – 3 are appropriate for Grades 4 – 12. However, Levels 4 – 6 are only suitable for Grades 7 – 12. All levels provide norms for comparison to student scores.

The test examples contain tonal patterns within the Tonal Concepts test and rhythm patterns within the Rhythm Concepts test. Gordon used a Moog synthesizer (Gordon, 1970, 1991) to produce the necessary audio samples. Gordon adheres to a gestalt (Mursell, 1937) ideal by utilizing real instrument sounds instead of pure tones (Seashore, 1915).

The ITML is hand-scored using scoring masks and normative scales. Each subtest has a different procedure for completing the musical examples in the Tonal Concepts test. For the Tonal Audiation/Listening test, students must decide the tonality of tonal patterns played on a recording. Next, on the Tonal Audiation/Reading test, they must determine whether the tonal patterns they see notated on the answer sheet are the same as the rhythm patterns they hear on the recording. Finally, in the Tonal Audiation/Writing test, students finish the notation on their answer sheet of the tonal patterns they hear in the recording.

There are also three subtests for the students to complete within the Rhythm Concepts test for each level. On the Rhythm Audiation/Listening test, students choose the meter of the rhythm patterns they hear. For the Rhythm Audiation/Reading test, students decide if the rhythm patterns they see notated on the answer sheet are the same as what they hear, and on the Rhythm Audiation/Writing test, the students finish the notation of the rhythm patterns they heard.

There are nine possible scores associated with the ITML: three separate scores for the subtests within Tonal Concepts and three within Rhythm Concepts, a Tonal composite score, a Rhythm composite score, and a whole-level composite score (Gordon, 1970, 1991). Combining individual section and composite scores represents the atomist and gestalt methodologies.

After using the Spearman-Brown Prophecy Formula on the subtests and the composite tests, the subtests were in the high .80s, and composites were in the low .90s for all grade-level groups (4-6, 7-9, 10-12) (Gordon, 1970, 1991). Therefore, the ITML demonstrated high reliability. Mohatt (1971) conducted a separate study to determine the validity of the ITML for his dissertation at the University of Iowa. The criterion-related validity coefficients were in the high .60s for the Tonal Total, the low .50s for the Rhythm Total, and the high .60s for the Tonal and Rhythm Composite. Therefore, the ITML demonstrated moderate criterion-related validity.

2.3 Subtests in Music Achievement Tests

Each subtest area included in this Literature Review reveals a student's possible achievement level in a specific area of musical development. Their descriptions are placed here in this Literature section to explain their nature and how they might function in a live assessment.

2.3.1 Subtests Often Used in Music Achievement Tests

2.3.1.1 Pitch Discrimination – Commonly used as a tonal test.

Duell & Andersen (1967) investigated the pitch discrimination performance of 168 1st, 2nd, and 3rd graders. The children assessed pairs of pure tones played on a tape recorder as either the ***"same"*** or ***"different"*** ("same" and "different" are also utilized in the MDAST). The beginning tone ranged from 390 to 440 cps, while the interval linking that tone and the comparison note ranged between 1/3 of a half-step and a major sixth. 68% of the children could discriminate **intervals as large as and larger than a half-step** (half-step intervals used in MDAST); however, 4% could not discriminate changes as great as a sixth. Performances improved from the 1st grade to the 3rd grade.

The Andrews & Madeira (1977) study tested the hypothesis that relational language ability may affect the ability to assess pitch discrimination. 36 "normal" children were divided equally into three age categories, six to six and 1/2 years, seven to seven and 1/2 years, and eight to eight and 1/2 years of age, and assigned five tasks. In task 1, a training procedure assessed the children's ability to hear the differences in the pitches of two tones, which were one octave apart. In task 2, the researchers evaluated the children's ability to label these pitches as high or low. Task 3 assessed the children's ability to compare two pitches and dictate whether the second

pitch was higher or lower than the first pitch. In task 4, the researchers examined the children's ability to label the position of a man on a ladder as high or low. In Task 5, they examined the children's ability to evaluate two men's positions on two ladders and discuss whether the second man was lower or higher than the first man. The results suggested that children who make pitch discriminations, like those demonstrated by nearly perfect scores on Task 1, often fail to exhibit those discriminations on assignments that require relational language. Comparing Tasks 2 and 3 to Tasks 4 and 5 indicates that children in the targeted age range are **less skilled in applying high-low and higher-lower to pitch than spatial relations** (Andrews & Madeira, 1977).

The ability to detect a shift of at least one semitone is known as pitch-change detection (Fancourt et al., 2013). As a result, one goal of this instrument will be for children aged 5 to 11 to detect variations of at least one semitone. Cooper (1994) found that students were substantially **better at detecting a shift in pitch than identifying the direction of the change**. Several authors stated that the pitch-direction test could have confused them because they used the verbal terms *higher* and *lower* or *up* and *down* for tonal directions (Cooper, 1994; Van Zee, 1976). Based on Andrews & Madeira's (1977) analysis, Fancourt et al. (2013) **advised against using the terms *high* and *low* concerning pitch**. Furthermore, their results indicate that a child's ability to discriminate pitch direction changes develops later in life than their ability to detect minor pitch changes (Fancourt et al., 2013).

Absolute pitch is an extraordinary type of pitch memory and long-term recall that some students with autism possess, affecting their results (Altgassen et al., 2005; Stanutz et al., 2014). Researchers found no evidence that cultural context affected the results of the tasks performed in a study of pitch perception and fundamental auditory discrimination among children from various cultural and musical backgrounds (White, Inuit, and Indian) (Walker, 1987).

2.3.1.2 Rhythmic Discrimination – Commonly used as a rhythmic test

Petzold (1963) conducted a five-year study in which he examined the progress of auditory perception in the subjects of "melodic perception, phrase learning, melodic reproduction with varying harmonies and timbres, and rhythmic ability" (Petzold, 1963, p. 26). The hypothesis he supplied was that the **participant's age** affects their auditory perception development. It was a significant factor, with some limitations. For most tasks, an **auditory perception plateau occurs by eight years old.** Moreover, indications corroborated that the most considerable development occurred between six and seven. These research findings (Zimmerman, 2007; p. 11) concluded that the "perception of musical stimuli follows a developmental sequence. Loudness discrimination develops first, with pitch and rhythm discrimination developing somewhat concurrently." Pitch and rhythm discrimination improve as the participant develops an increasing attention span and memory improvement. A child inherently acquires loudness perception without conventional training. Therefore, **pitch and rhythm discrimination can become the foci of instruction.**

Zimmerman's review of research findings relating to children's musical characteristics establishes a sequential progression for their development. According to Zimmerman, there is some indication that in vocal growth and rhythm discrimination, the development relies more on the **participant's maturity level** than on a certain kind of instructional or environmental experience.

2.3.2 Rationale for Subtests To Be Used in Music Achievement

2.3.2.1 Musical Contour Identification – Sometimes used as a tonal extension.

In the Fyk (1995) study, they conducted two experiments to (1) verify the effect of the contour type, pitch range, and length of melody on the observability of melodic contour by children and (2) assess their capacities to perceive pitch changes in melodies with altered contours. For Experiment 1, the researchers examined 60 subjects: five- and six-year-olds. These children listened to five-tone and ten-tone melodies with different contours and pitch ranges. Their assignment was to identify the kind of contour. **Descending,** rather than ascending, **contours were easier for the children to recognize**. In addition, the children **identified ten-tone rather than five-tone melodies more consistently**. In Experiment 2, the assignment for 30 five-year-olds was to identify a change in the melody's second tone. When an interval of a **major third** altered the second pitch, there was **50% accuracy** for the alteration and **82% correct** when the alteration exceeded an **octave** (Fyk, 1995).

In the Pick et al. (1988) article, they investigated children's perception of scale and contour in melodies in five studies. Experimental assignments included assessing transposed interpretations of melodies (Studies 1 and 3), discriminating between transposed interpretations of a melody (Study 2), assessing contour-preserving alterations of melodies (Study 4), and assessing similarity to a familiar focal melody of alterations preserving rhythm, or rhythm and contour (Study 5). The first and second studies showed that young children could detect key transposition alterations in familiar melodies and recognize similarities over key transpositions in unfamiliar melodies. Young children are also sensitive to melodic contour over alterations that preserve it (Study 5). However, they still instinctively distinguish between melodies that maintain the same contour and have different intervals (Study 4).

2.3.2.2 Musical Composer Comparison – A new subtest in the Literature.

According to Marshall and Hargreaves (2007, p. 33), "Gardner (1973, p. 326) operationalized style sensitivity in music as the ability to recognize whether two contrasting musical pairs came from the same or a different piece of music, that is, as 'the ability to group works produced by one artist.'" While Gardner initially discussed musical style, he supported another important musical ability with this comment. If a child can group works according to whether or not a single artist (or composer) produces them, they can differentiate when works are not by the same artist (or composer). Children can compare and contrast pieces of music against one another and make a value judgment regarding their status. Therefore, the child can answer the following question: Are the pieces of music composed by the same person, or are they written by a different person? An Ilari and Polka (2006) study indicated that 8-month-olds could discriminate between two similar pieces of music (in meter and tonal center) by Ravel from his *Le tombeau de Couperin*: Prelude and Pavane. If infants can discriminate between whether pieces of music are by the same composer, older children should also be able to do so.

2.3.2.3 Musical Styles Comparison - A new subtest in the Literature.

Gardner (1973) found that first graders had a positive style response and that the mean scores at the upper age levels were high in his musical style study. He partially credited the success of his task's performance to the gifted target population he had selected for the report. He speculated that a group of less intelligent or motivated children might have produced different results (Gardner, 1973). However, his works bear out that he also thought that "an important consideration in perceiving styles is whether two works feel differently or have distinct Gestalten," he suggested (Gardner, 1973, p. 75). Therefore, he did not attribute the

ability to perceive musical styles as dependent on a student's level of intelligence so much as the ability to recognize whether two pieces feel the same or different.

In addition, Marshall and Hargreaves (2007) suggest that many preschool children seemed capable of perceiving differences in musical styles. They credited their study's success with Gardner's success in his pursuit of sensitivity to styles within music and the arts. Marshall and Hargreaves described his study as "pioneering research" (Marshall & Hargreaves, 2007, p. 33). They provided a synopsis of his study, which he conducted with children's interpretations and understanding of musical style.

The children in the study ranged from age six through eighteen. First, they selected Western classical musical examples from 1680 to 1960. Then, Gardner had the participants play pairs of short musical excerpts that exemplified four musical periods: Baroque, Classical, Romantic, and Modern. (A side note here is that he equated musical style with a musical era. Thus, this illustrates the crossover between style and era I wanted to avoid in my study. They were too highly correlated.) Gardner's experiment included a systematic comparison of periodic styles against one another. For instance, he used Baroque with Classical, Romantic, and Modern, then Classical with Baroque, Romantic, and Modern. After listening to a set of excerpt pairs, the participants were asked to choose whether they believed the two excerpts were from the same or different pieces of music. The results revealed that even the youngest participants in the study (age six) demonstrated high sensitivity to musical style.

Additionally, in a study of musical style preferences and aural discrimination skills, May (1985) suggested a positive relationship between primary school children's music preferences and their abilities to make same/different decisions about musical stimuli pairs (Van Zee, 1976).

Therefore, the May and Van Zee studies were positive continuations of the literature regarding children's ability to make musical style-based decisions.

2.4 Summary of the Literature

This literature review aimed to prepare the researcher with a foundation of knowledge about the subject area of interest (MDAST contributory information) before they embarked on the study that would follow it. By the time I began my Literature Review, I had already crossed off a short list of research requirements for this study. Therefore, I brainstormed and drew a conceptual map to create a structure for the research I needed.

I filled in the topics/headings logically and assigned them to populate the Chapter 2 area of the Table of Contents. That gave me the essential subject areas that unfolded in this section. 1) The chapter opened with (2.1), a comparison between aptitude and achievement, which was a pivotal discussion, especially since I initially intended the MDAST to be an aptitude test, had a switch to achievement, and now a final realization that the MDAST is both an aptitude and achievement assessment. This research gave a deeper explanation to support my rationale in "Chapter 1" about my determination that this MDAST test was an aptitude/achievement test. 2) The second section (2.2) introduced and compared three respected music achievement measures and their creators (Aliferis, 1957; Colwell, 1970; and Gordon, 1970). The sections that followed the achievement measures introduced Subtests in Music Achievement Tests and a rationale for using them.

Chapter 3: Method

3.1 Inspiration to Create

I was used to creatively assembling assessments during my tenure as a principal and educator. For 16 years, I owned a school of the arts for special needs children, Center for Education School of the Arts and Sciences, in Tampa, Florida, and developed the curriculum from the bottom up. Therefore, when the time came to create an assessment for a musical measurement class at the University of South Florida, I **thought** I was somewhat experienced and prepared for the challenge. As a result, my new measure was born.

I gave my assessment two rules that I knew must be fulfilled for this study to be successful, and I stated them here again because this is the consistent direction my dissertation was supposed to go. The **first** purpose of this dissertation instrument was to document the development of a new music achievement test, the Musical Discrimination and Styles Task (MDAST). The **Final** purpose described the strength of the evidence supporting the validity and reliability of this new developmentally appropriate music aptitude and achievement instrument.

3.2 Versions of the Measure

The MDAST began as a simple idea but it multiplied quickly after its inception. It was originally an assignment for a Measurement and Evaluation in Music class, which required the creation of a moderate measure for an acceptable grade. In its infancy, the MDAST started as my modest attempt at a music aptitude test that was small and straightforward. However, I did not wish to recreate or imitate “pitch and rhythm” versions of aptitude measures already in existence

(Gordon, 1989). Therefore, I expanded the parameters of the test in an attempt to determine whether other musical subject areas were inherently observable in a child's natural development. The original MDAST (now known as the MDAST Long Version) was used in the pilot study and contained five subtests: (1) pitch discrimination and several comparison sections between (2) musical contours, (3) musical composers, (4) musical styles, and (5) musical eras. These subtests contained ten questions and two practice samples each. For each example in the subtests, students listened to pairs of short musical notes or phrases and judged whether they sounded the same, different, or "I don't know." There was only one correct answer for each question. The student population was K – 5, and the total test time lasted forty-five minutes to one hour.

3.3 A Brief Pilot Study Overview

For the pilot study, the purpose of the MDAST was to determine the *music aptitude* of elementary music students in Grades K – 5 (ages 5 – 11). Therefore, it subdivided into five subtests of comparisons: pitch, contours, composers, musical styles, and musical eras. Each section included ten questions and two practice questions. The response format was "same, different, or 'I don't know.'" The last response was to prevent students from guessing on the test. I took the sample from a Montessori School in Florida. The study occurred from October 12 to November 22, 2020. I planned the methodology in phases. First, I reviewed the literature and established a theoretical framework. Second, I created the instrument and a corresponding website. Third, I had a panel of experts with terminal degrees in music review the items on the MDAST to 80% agreement. Next, I had the student sample take the test. Finally, I conducted the necessary statistical analyses. I ended up with a small sample size (n=7) (the study occurred during the COVID pandemic's height), which impacted the results.

Unfortunately, the analysis excluded many items due to the lack of variance from a small sample size. After this, only 27 of the 50 items remained, with an initial Kuder-Richardson Formula-20 of .434. [Table 1- p. 39] Therefore, I removed seven items based on the item-total results. Consequently, removing those items led to an overall KR-20 (Cronbach's alpha) of .854. After removing those seven items, the best outcome was that all other items bumped into the .800s. [Table 2- p.40] That led me to realize that those seven items needed to be replaced and with new evaluations by a panel of experts to determine if they were acceptable items to add to the MDAST assessment.

I was careful to make adjustments that would balance the subtests. Eventually, I would create a different order for the subtests: pitch discrimination, rhythm discrimination, contour comparison, composer comparison, and musical style comparison. I eliminated the musical eras subtest because it had more mistakes than any other portion of the test. Musical styles and eras are also closely related and can be almost interchangeable (Gardner, 1973). The eras subtest removal was a logical step.

Table 1.

*PILOT: Kuder-Richardson Formula 20: Item-Total Statistics (**Before Adjustments**)*

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
No_9	18.7143	8.905	.401	.400
No_12	18.7143	8.905	.401	.400
No_18	18.8571	8.143	.564	.353
No_19	19.0000	9.333	.102	.446
No_20	18.7143	10.571	-.310	.501
No_21	18.7143	8.905	.401	.400
No_22	18.7143	9.238	.249	.424
No_24	18.8571	7.810	.698	.323
No_28	19.0000	8.667	.318	.400
No_29	18.8571	8.143	.564	.353
No_31	19.2857	8.571	.400	.387
No_32	18.8571	10.143	-.138	.488
No_33	18.8571	7.810	.698	.323
No_34	18.7143	8.905	.401	.400
No_35	18.8571	10.143	-.138	.488
No_36	18.7143	10.571	-.310	.501
No_37	18.8571	8.143	.564	.353
No_40	18.8571	10.143	-.138	.488
No_43	18.7143	9.238	.249	.424
No_44	18.8571	11.810	-.625	.566
No_50	18.7143	10.571	-.310	.501
No_54	19.0000	11.000	-.376	.536
No_55	18.8571	11.143	-.439	.538
No_57	19.0000	12.333	-.710	.590
No_58	18.7143	8.905	.401	.400
No_59	18.7143	8.905	.401	.400
No_60	19.1429	7.476	.749	.298

Note. The Kuder-Richardson Formula 20 is related to Cronbach's Alpha. The results from KR20 analyses reflect the findings as a Cronbach's alpha values in SPSS. Items in boxes were removed from the final MDAST to improve reliability.

Table 2.

*PILOT: Kuder-Richardson Formula 20: Item-Total Statistics (**After Adjustments**)*

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
No_9	13.5714	20.286	.448	.847
No_12	13.5714	20.286	.448	.847
No_18	13.7143	19.238	.578	.841
No_19	13.8571	19.810	.390	.850
No_21	13.5714	20.286	.448	.847
No_22	13.5714	20.952	.248	.854
No_24	13.7143	18.571	.747	.834
No_28	13.8571	20.143	.318	.853
No_29	13.7143	19.238	.578	.841
No_31	14.1429	18.810	.686	.836
No_32	13.7143	20.905	.181	.858
No_33	13.7143	18.571	.747	.834
No_34	13.5714	20.286	.448	.847
No_35	13.7143	21.905	-.042	.867
No_37	13.7143	19.238	.578	.841
No_40	13.7143	20.905	.181	.858
No_43	13.5714	20.952	.248	.854
No_58	13.5714	20.286	.448	.847
No_59	13.5714	20.286	.448	.847
No_60	14.0000	18.333	.728	.833

Note. Observe the highly significant relationships in the Cronbach's alpha column since the previous items in Table 3 were removed.

A panel of experts with terminal degrees in music (n = 5) had previously assessed and verified the content validity. But, based on the other results, the reliability was limited due to the sample size and exclusions. During the expert analysis phase, the eras subtest had more faulty items than anywhere else in the test. Therefore, when I decided to add a Rhythm subtest to the measure, I removed the Eras subtest based on the Era's faulty items.

However, since this pilot study, I've learned that removing items from the MDAST in this fashion might not have been a good choice. I only had an (n = 7), which didn't allow

accurate studies to run on the assessment. Instead of making all those changes, I could have chosen another option: to leave everything as it was and try to retest with a larger population of students to get a more accurate picture of the assessment's validity and reliability.

The pilot study concluded with a verified content validity and reliability of .854. However, based on these results, a more extensive study was warranted to confirm the results on a larger scale. I chose to pursue future research on this project because this test was the only one to address all five essential content areas in children simultaneously. Specifically, they were (pitch discrimination, [rhythm discrimination—later added], musical contours, musical composers, musical styles, and musical eras [later cut]). Additionally, I believed there could be future uses for the MDAST, which included extending research within the music education field.

3.4 The Four Phases of Initial MDAST Study's Research and Development

The online assessment designed for this study went through a four-phased research and development approach. First, I reviewed the literature on musical aptitude (which, since then, amended to reflect both aptitude and achievement), the leaders in the field of music aptitude and achievement, the conventional items found on a musical aptitude or achievement test, and a discussion of aptitude versus achievement. The review led to the development of a theoretical framework for the MDAST. During phase one, I developed the initial survey. In phase two, a panel of experts (n=5) with terminal degrees in music reviewed the instrument, and I conducted item analysis based on their responses. A student sample group interacted with the instrument in phase three. Finally, I determined the survey instrument's validity and statistical reliability analyses in phase four. Throughout each phase, findings informed the primary researcher to refine the survey measure systematically. The goal was two-fold: to contribute a rigorously

reliable and valid aptitude instrument to the field of music education and to prepare for future studies using the MDAST.

3.4.1 Phase One (Pilot Study): Test Development

I developed the initial survey to determine if children aged 5 – 11 could differentiate these concepts: pitch discrimination, musical contours, composers, musical styles, and musical eras. These concepts formed the five subtests of the instrument. Each subtest consisted of 10 questions with two practice questions. I developed each question to contain two parts: one functioned as a "question" and a second function as an "answer." The student had to determine whether the question and answer sounded the same, different, or "I don't know."

Because Fancourt et al. (2013) cautioned about using the terminology *high* and *low* concerning pitch and contour (Andrews & Madeira, 1977; Cooper, 1994), an alternative way of describing the aural stimuli had to be developed. This concrete reasoning was behind the "same, different, and I don't know" format. Additionally, based on the literature review, I determined to keep all discrimination interval items within one semitone (Fancourt et al., 2013).

In the MDAST Long Version, items 1 – 25 (Sections one and two- omitting #15) were musically composed by the primary investigator for inclusion in the MDAST. I recorded all items individually with a sine tone generator and an NCH Tone Generator and sequenced them with a WavePad Sound Editor. Sections three through five (items 26 – 61) consisted of pairs of musical excerpts taken from YouTube. They were converted via cellular phone to mp3 format with Tube Media Downloader and edited for 15-second size with the Cut Ringtone Cut Music app. Due to the limited duration of each sample and their applications under educational fair use,

the primary investigator felt secure utilizing this method. The list of available items on the MDAST (Long Version) is in [Appendix B].

The researcher assembled the online assessment on the Qualtrics Survey Platform. I uploaded all audio samples into the system. Next, the primary investigator wrote items corresponding to the audio samples and the appropriate subtest to which they belonged [Appendix C]. Voice recordings of the items were also created and uploaded for younger students who may have difficulty reading some words. The order within each item consisted, first, of a push button for the voice recording of the written directions; second, of a written version of the directions; third, of a push button for an audio sample of the "question" example, and finally, of a push button for an audio sample of the "answer" example. I carefully crafted each item to have only one correct answer for each question. The MDAST took approximately 45 minutes to one hour for a participant to complete in its finished state.

Once completed, I linked it to a website specifically constructed and designed on a WIX.com platform for the MDAST. At www.mdast.online, potential families could interact with the primary researcher (me). Parents could book appointments to speak with me. They could also call me through the website info. It contained a copy of the Parent Verbal Consent and the Child Verbal Assent scripts so families could follow the document with me in consent discussions for participation in the study. Once a parent had given their consent and the child participant had given their assent, I gave the family a password to the MDAST. I embedded the link to the MDAST directly on the website.

3.4.2 Phase Two (Pilot Study): Panel of Experts

I included the music faculty and doctoral music students from the University of South Florida Center for Music Education Research in the School of Music in developing the MDAST measure. This approach is the most common one to ensure content and face validity, whereby a panel of experts reviews the items for completeness, logic, and clarity (Trochim, 2006). A combination of five music faculty and doctoral students took the MDAST before making it available to any student population. I removed items in which less than four members (80%) agreed to the correct answer from the measure and replaced them with a new item.

3.4.3 Phase Three (Pilot Study): Student Sample Group (n = 7)

I was granted three advertisement flyer distributions at the Montessori School in Florida to promote student interest. Unfortunately, despite these three attempts, only six families responded (seven students total) from a student population of seventy-five at the school. The lack of overall participation in the study was disappointing, but it was understandable, considering that it occurred during the height of a worldwide pandemic. Therefore, the available sample data was recorded and analyzed.

3.4.4 Phase Four (Pilot Study): Statistical Analysis

In a quest to establish internal reliability, I performed a Kuder-Richardson Formula 20 analysis. Researchers use the KR20 with tests, like the MDAST, that deal with absolutes, like "true and false" or "same and different," scored as "right or wrong." I conducted bivariate correlations on the individual subtests to determine if they have a significant relationship.

3.5 The Pilot Study Continues

This section of the dissertation explains how I conducted the pilot study with the original MDAST, now known as the MDAST Long Version.

3.5.1 Participants and Sampling

Volunteers came from the Montessori School in Florida, between October 12, 2020, and November 22, 2020. While there was no formal music education program, the after-school program offered private violin lessons to the student population (K – 8). Only participants who completed a phone interview with the primary investigator, completed a verbal assent, and had a parent complete an *Informed Verbal Consent to Participate* were allowed access to the online survey.

3.5.2 Demographic Information

Four boys and three girls from Community Montessori School participated in the MDAST study. Their ages varied from kindergarten to grade 4. The participants were female (42.8 %) and male (57.2%). The mean age of the sample was 7.29 years (SD = 1.22).

3.5.3 Procedure and Data Collection

Teachers at the Community Montessori School sent recruitment flyers for the MDAST study home with all students in the school. Interested parents could call me (the researcher) or book a meeting on the study's website. Once an online or phone meeting began, I reviewed the Verbal Informed Parental Consent and explained the study's procedures. If a parent agreed for their child to participate, then I discussed a Child Assent form with the child. If the child decided to participate, I gave them the password to enter the study. The porthole resided on the website.

Once the child entered the password, the MDAST began. Students took the test on a computer, tablet, or phone.

As a child participated in the MDAST, all the answers were recorded and graded by the host site, Qualtrics. They also housed basic statistical information on the test results, such as chi-squared. However, I did not need those statistics at that time. Therefore, I did all my statistics in SPSS 27 for the pilot study.

3.5.4 Data Analysis

3.5.4.1 Content Validity

When the USF music faculty and doctoral music students (*n*=5) took the MDAST, the intention was to establish content validity. I removed items on the test with greater than one dissent regarding the correct answer (4/5 or 80%) from the task and replaced them with new ones. I replaced nine items after the faculty/doctoral review. They included items 19, 28, 31, 35, 44, 54, 55, 57, and 60. Following the deletions and replacements, I verified the content validity.

3.5.4.2 Reliability

I pursued internal reliability by utilizing interitem analyses via the Kuder-Richardson Formula 20. The initial analysis excluded several items because their variance was undetectable with such a low sample. Therefore, only 27 items remained available for discernment from the original data. The KR20 is a variant of Cronbach's alpha analysis, so it is performed similarly in SPSS. In essence, the output of the analysis received from the KR20 is Cronbach's alpha value. The Cronbach's alpha for the first analysis of the MDAST was .452. By studying the item-total

statistics, I determined that the removal of seven items would result in the improvement of the overall reliability [Table 2]. In response, I removed items 20, 36, 44, 50, 54, 55, and 57.

I repeated the KR20 interitem analysis with the 20 remaining items. At this point, Cronbach's alpha for the second analysis yielded .854. Additionally, a study of the item-total statistics [Table 2] revealed that the overall reliability of the individual items had significantly improved into the mid and high .80s.

3.6 Modifications Based on Feedback

On September 8, 2021, in the Center for Music Education Research Seminar attended by approximately ten doctoral students and three faculty members in music education, I delivered a PowerPoint presentation designed to describe and explain the MDAST in preparation for a dissertation proposal. The presentation was well-received, with most information accepted as a viable and cohesive starting point for writing the proposal. After the presentation, the faculty and my colleagues held a brainstorming session to provide feedback. They surmised the study elements were working positively and expressed which items may require modifications or should be discontinued.

At the outset of the discussion, one of the biggest concerns was whether the MDAST was a test of music *aptitude* or music *achievement*. An aptitude test measures a participant's natural ability to do something or its inherent or latent affinity for success in a given subject area (Snow, 1991). However, an achievement test is an ability that was already done successfully and manifested via effort, skill, or courage on material already learned (Kaur et al., 2016). The discussion focused on the last subtests of the MDAST: musical composers, musical styles, and musical eras. Most seminar participants in the discussion felt that a person only attained these

three subtest areas once a person experienced familiarization with them through musical introduction and instruction. Based on this discussion, I decided to undertake research that compares and contrasts aptitude and achievement tests to make an informed choice regarding the MDAST's identity. The research on this subject is included in the literature review of this document.

I addressed this question earlier in this dissertation, but I feel it is essential that I reiterate why I assert that this test qualifies as ***both an aptitude and an achievement test.*** While the first two subtests could be used on a traditional aptitude test (pitch and rhythm), there are also three subtests (contour, composers, and style—in the *MDAST Abbreviated*) that could be used on an achievement test. They test the amount of retention from a participant's repeated exposure to a musical element in their environment or culture. The participant grows to recognize musical items and gradually makes comparisons between items they have heard (Gordon 1990). Therefore, after thoroughly researching this subject, and pondering this "aptitude or achievement" question for three years, I believe that this test qualifies as both a musical aptitude and achievement assessment.

On September 18, 2021, I was honored to deliver a poster presentation at the Asian Pacific Symposium of Music Education Research (APSMER) entitled "*Musical Discrimination and Styles Task: A new possibility for assessment of music aptitude in elementary music students*." In the breakout room, a robust discussion questioned why I had not included a rhythm subtest in the MDAST. While the measure, purpose, and statistics elicited positive feedback, the group agreed that the lack of a rhythm subtest was a fundamental design flaw. I agreed with the consensus among the participants in the discussion and determined that I would add a rhythm

subtest to the test battery before the subsequent study occurred. Since then, I have added a rhythm discrimination subtest to the assessment.

In summary, I originally wanted to create an aptitude test. However, as it turned out, some of the data I hoped to glean from participants taking the measure was incompatible with an aptitude test (the last three subtests). I included the tonal and rhythm subtests that were familiar to one. But then, I also had contour, composer, and style comparison subtests. These subtests could only measure a student's knowledge level after learning. I had not based them on a student's inherent natural abilities.

Based on this reasoning, research, and feedback from my professors and colleagues, I determined this measure could not be an aptitude assessment like one I originally wanted to create. Neither could it be a purely designed achievement battery like I thought I was settling for. I originally wanted to create something new that pushed the boundaries of aptitude tests that were already in existence. I realized that I couldn't find a way to fit either of those definitions because I created something new. The MDAST fits the profile of both a musical aptitude and achievement test.

I spent a great amount of time trying to justify why MDAST was an aptitude test, and then just as much time (probably more!) trying to explain how it was an achievement assessment. As I took the time necessary to ponder whether creating another achievement test was essential in music education, I researched the standard musical achievement tests currently in use. I was surprised to learn that almost everyone looking for achievement tests harkened back to studies performed in the 1950s, 1960s, and 1970s (Aliferis, 1957b; Colwell, 1970; Colwell, 1969; Gordon, n.d.). Some of these innovators also created music aptitude tests, and my line between aptitude and achievement blurred. I understand why I felt like the line between aptitude and

achievement was blurred about the MDAST—because I kept asserting that the assessment had to be one thing or the other (aptitude or achievement). I would not allow myself to consider the possibility of both. However, the more I read about how Richard Colwell and Edwin Gordon distinguished aptitude and achievement from one another, the easier it became for me to understand. I had already researched their theories about music aptitude (Gordon, 1989), achievement (Aliferis, 1957b; Colwell, 1969; Gallagher, 1971; Gordon, n.d.), and audiation (Gordon, 1999) and included some of them in my theoretical framework. As I adjusted my focus from aptitude to achievement to both, I realized that I could take the best of their research and examples and move in a new direction to create a new paradigm. Therefore, I adapted my test for a combined purpose and new audience.

3.7 Discussion of Pilot Study

At that time, the pilot study was still categorized as a music aptitude measure and was initially meant to encapsulate pitch discrimination, melodic contours, and the differentiation of composers, styles, and musical eras. The instrument's content validity was assessed through a process of evaluation by a panel of five experts with graduate degrees in music. After removing and replacing items that displayed less than an 80% consensus of the correct answer, I verified the instrument as valid.

However, despite its high Cronbach's alpha results, the MDAST had limited internal reliability due to its small sample size. In addition, due to the number of items excluded from the initial analyses, a more accurate picture of all items on the instrument was impossible. Therefore, a future study with a larger sample was warranted to determine the true internal reliability of the MDAST.

3.8 Dissertation Preparation: Modifications Since the Initial Study

Once the pilot study ended, preparation for the dissertation study began. Since its creation, I attempted to adapt the MDAST to the student population it serves. Therefore, I tried to find ways to make the test flexible for different groups and situations. The ability to administer the assessment individually or to groups (as individual logins in a computer lab) was one example of the flexibility I'd hoped for.

However, I learned from the pilot version that it took 45 minutes to an hour for elementary students to take the *MDAST Long Version*, which was not what I had initially intended. After the participants struggled to remain focused for such a long assessment, I determined that any MDAST version that took 45 minutes to an hour to complete would be better suited for high school and college students. I'm not sure if the difference in test length will make a difference in student outcomes within the results or if musical experiences (and opportunities to make music on instruments) will be primary reasons for music achievement. I will conduct a future study to determine whether the *MDAST Long Version* is appropriate for these groups of students and the reasons for the achievement.

After the experiences with the MDAST Long Version, I created a version with five subtests of five questions each and one practice question per section. The timing of this version of the test was 36 minutes. Once again, I determined that this version of the MDAST was too long for elementary-level students. I renamed this version the *MDAST Short Version*. Moving forward, I could eventually do a study(s) to test its effectiveness in music achievement in children ages 12 – 15 (in the United States, Grades 7 - 9). I would also need to determine whether the achievement level varied between studies in different grade levels (the levels should gradually increase by grade level). I could test students in the music classes but tried to test at

least three non-music courses (one for each grade level) to serve as control classes. I also wanted to test to see which music classes would get the highest achievement level and if the MDAST could perform that way. If the MDAST could perform this way, it would be helpful for music directors in the classroom. They could use it to determine whether their students needed help in a specific area in music class.

Finally, I shortened the test one last time. As the final iteration, the *Musical Discrimination and Styles Task Abbreviated Version*" is also the test of music achievement used as the basis of this study. I intended this test for children aged 6 – 11 (U.S. grades 1 – 6). It officially consists of five subtests of three questions and a practice question each, and it currently takes approximately 20 minutes for each participant to finish. In addition, I included a sixth section of four survey questions relating to test performance in the instrument. By creating three (3) different versions for three (3) different age groups, I felt that the MDAST was meeting another flexibility component I'd hoped for—being available for all grade levels.

It is important to note that I changed which subtests remained on the test for all MDAST versions. The five current subtests (in each version) are comparisons in the following categories: pitch discrimination, rhythmic discrimination, musical contours, musical composers, and musical styles. The musical eras subtest was removed from the test because it correlated too highly with the musical styles subtest. It was also the subtest with the most errors in the Pilot Study.

3.9 Dissertation: Study Order of Research Methods

In this proposal section, the methodological thought and processes in my research plan, the "Order of the MDAST Research Design," seen in [Figure 4] was how I originally planned to tackle my research. However, [Table 3] is how my research design unfolded as I allowed my

original **first** and **second** purposes from "Chapter 1" to help guide what I needed to know from my study. Those refinements led me to adjust my research questions, and I began digging into the necessary analyses to answer them. Inevitably, that required me to change my plan and my research design order.

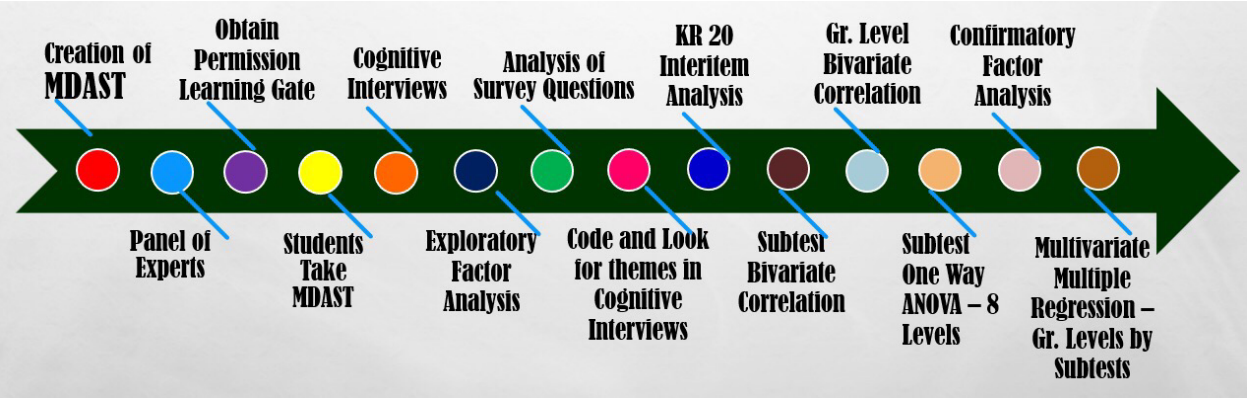

Figure 4. *Previous Order of MDAST Research Design*

Table 3.

Dissertation: Final Order of MDAST Research Design

Item	Procedure
1	Creation of the MDAST
2	Review by a Panel of Experts with Terminal Degrees in Music Education (N = 5)
3	Obtain permission from the Community School
4	Obtain approval from the Institutional Review Board
5	Participants take the MDAST (N = 357)
6	Descriptive Statistics: Item Difficulties and Correlations of Subtests
7	MDAST: Expert Item Analysis (Content Validity)
8	Confirmatory Factor Analysis (Internal Structure Validity)
9	K.R. – 20 Inter-item and Subtest Analyses (Reliability)

3.10 Steps in Developing the MDAST

The ten steps used to develop the MDAST match the questionnaire construction process by Crocker and Algina (1986). First, I identified the purpose of the instrument. The instrument was intended to be used as both a research and an educational tool. Second, I confirmed that no other existing measures would satisfactorily serve the purpose. In the third step, I defined the theoretical constructs and content domains. The content consisted of five subtests: pitch, rhythm, contours, composers, and styles. Next, I generated a preliminary item pool for these five subtests. For the sixth step, I submitted the measure to a panel of five experts with terminal degrees in music. I asked the experts to take the test, and for an item to be included, 80% of the experts had to agree on the correct answer. Then, I pretested the measure with a preliminary tryout of seven

students in grades one through five. Next, I conducted additional analyses with a sample of 357 participants, including item analysis, confirmatory factor analysis, and reliability analysis, to determine how well the items were functioning. Then, I created procedures for administering, scoring, and interpreting the scores in the next step.

3.11 Population and Sample

The 357 participants came from one charter school in the southeastern part of the United States. To be eligible for this study, participants needed to have informed parental consent and child assent. Students consisted of students in grades three (n = 85), four (n = 89), five (n = 100), and six (n = 83). Most students (75%) confirmed that they participated in music classes at the school. Most students (83%) did not take private lessons. The decision to use over 300 participants was based on a power analysis in which I was interested in detecting an effect between the student's grade level and the student's scores on the five subtests. The total number of participants (n=357) exceeded the required 260 participants needed to achieve power requirements, based on calculations from G*Power. I based that number on an effect size of 0.35, an alpha of 0.05, and a power of 0.80. I chose that effect size level based on ***moderately conservative*** effect sizes found in *similar literature* on musical measurement instruments (Gordon, 1989). The study contained 362 participants to account for attrition (children who did not finish the measure) since that was a severe detriment to the pilot study's success. This time, I attempted to meet the power threshold by aiming for a higher sample size. Additionally, more participants allowed me to conduct a viable confirmatory factor analysis.

3.12 Issues of Diversity

The MDAST is only available in English at this time. However, in a future pilot study, I plan to have the instrument translated into Spanish to determine its validity and reliability and establish a more equitable and diverse testing opportunity for students.

3.13 Data Collection

Having received written approval from the University of South Florida's Institutional Review Board, I emailed a research package to the school principal. It contained the necessary parental consent forms, informed consent forms (for teachers), and the web link and password to enter the MDAST. In addition, parents and teachers could view a mini-sample version of the test. I personally supervised and administered the data collection for 6th-grade students at the Community School in their computer lab. The 3rd – 5th grade teachers at the Community School administered the data collection phase of the study in the school's computer lab. Qualtrics, which is a survey platform, houses the MDAST instrument. Students who received signed parental consent came to the computer lab to log into the measure and take it online. The MDAST instrument includes the child's assent in its introductory portion. If the child agreed to participate, they continued into the instrument. If they declined, the test redirected them out of the instrument. Once a child took the assessment, Qualtrics automatically graded it and saved the results for future analysis. On average, students took 20.7 minutes to take the test (median = 18.3). Time to complete the test ranged from 6.1 to 201.7 minutes.

Chapter 4: Results

The second purpose of this dissertation is to describe the strength of evidence supporting the various kinds of validity and reliability of the *Musical Discrimination and Styles Task* (MDAST). "Chapter 4: Results" contains the findings of all the analyses conducted in the dissertation version of the MDAST Abbreviated study.

4.1 Introduction

As I embarked on my "Results" chapter, I knew that the purposes I established at the beginning of this dissertation were essential to its success. I had used them to keep me on track thus far, and they worked as I intended. The **first** purpose was to document the development of a new music aptitude and achievement instrument entitled the Musical Discrimination and Styles Task (MDAST), and the **second** purpose of this dissertation was to describe the strength of evidence supporting the validity and reliability of this developmentally appropriate music achievement instrument for elementary and middle school music students. However, as the study progressed to "Chapter 4: Results," I further defined those purposes because they (especially the second one) would have such a profound effect on the study. The **first** purpose recorded the MDAST's creation. It included two subtests that assessed students' aptitude to discern pitch and rhythmic discrimination. However, there were also three achievement subtests for the students' knowledge of musical contours, the ability to differentiate between composers' works (or identify the same composer's works), and the ability to recognize musical styles. The **second**

purpose of the study pertained to the MDAST in action as students took the assessment, and the focus shifted to the determination on the levels of the MDAST's validity and reliability.

Therefore, I intended to provide a rigorous connection between the research's planning and results for Chapter 4. The proposal listed the prior study performed on the MDAST and revealed how the results unfolded after the Pilot Study. There were lessons learned and some pitfalls, too.

Nevertheless, the dissertation study was strikingly different by design, so it was natural to begin the analysis with comparison and contrast observations. The Pilot Study (PS) only had an (n=7), while the Dissertation Study (DS) boasted (n=362). Even though there were similarities between the two studies (e.g., I selected all the test items from the original 300-item bank), there were more differences between how I delivered the tests to the students, how I obtained the parent consent and student assent, how I interacted in the same room with the students (NO COVID), and a different school served as the student population.

After updating the MDAST and my first study's requirements to be successful with elementary students and receiving written approval from the University of South Florida's Institutional Review Board, I administered the *MDAST Abbreviated Version* at the Community School in their computer lab. Qualtrics automatically graded it and saved the results for analysis. I revealed the Method and charts of statistics in "Chapter Three: Methods." "Chapter Four: Results" was organized into four essential sections (4.1) Introduction, (4.2) Data Description, (4.3) Research Question Results, and (4.4) Summary.

"Chapter 4.1", the Introduction section, revealed common information that connected the "Chapter 3: Methods" section with "Chapter 4: Results." It also prepared the researcher and subsequent readers for the structure of the chapter. The Data Description section (4.2) described

the collected data. It presented the item difficulties for the 15 items in the MDAST for both the total group and by grade level. In the (4.3) "Research Question Results" section, I used the study's findings to answer the research questions. I provided each research question within this chapter, briefly portrayed the procedure, and presented the results. Finally, in (4.4) "Summary," I summarized the results of the analyses presented within the chapter.

4.2 Data Description

Table 4 presents the item difficulties for the 15 items in the MDAST for the total group and by grade level. Item difficulties for the total sample ranged from .22 (Contours Q19) to .95 (Composers Q38). For the **total sample**, the hardest items (item difficulties < .50) were .22 (Contours Q19), .27 (Composers Q32), .29 (Styles Q44), .31 (Composers Q30), and .47 (Contours Q17).

Item difficulties for the **third-grade** sample ranged from .15 (Styles Q44) to .94 (Composers Q38). For the total sample, the hardest items (item difficulties < .50) were .15 (Styles Q44), .16 (Composers Q32), .21 (Contours Q19), .32 (Composers Q30), .41 (Contours Q17), and .45 (Styles Q41).

Item difficulties for the **fourth-grade** sample ranged from .20 (Contours Q19 and Styles Q44) to .98 (Composers Q38). For the total sample, the hardest items (item difficulties < .50) were .20 (Contours Q19 and Styles Q44), .22 (Composers Q32), .24 (Composers Q30), and .39 (Contours Q17).

Item difficulties for the **fifth-grade** sample ranged from .16 (Contours Q19) to .96 (Contours Q20). For the total sample, the hardest items (item difficulties < .50) were .16 (Contours Q19), .37 (Composers Q30), .39 (Styles Q44), and .49 (Contours Q17).

Item difficulties for the **sixth-grade** sample ranged from .29 (Composers Q30) to .93 (Composers Q38). For the total sample, the hardest items (item difficulties < .50) were .29 (Composers Q30), .34 (Contours Q19), and .36 (Composers Q32).

Table 4.

Item Difficulties by Total Group and by Grade Levels

	Total (*n*=357)	Grade 3 (*n*=85)	Grade 4 (*n*=89)	Grade 5 (*n*=100)	Grade 6 (*n*=83)
Pitch Q5	.81	.72	.81	.91	.78
Pitch Q6	.75	.71	.66	.86	.73
Pitch Q7	.86	.78	.85	.92	.88
Rhythm Q176	.79	.71	.79	.83	.82
Rhythm Q177	.85	.84	.87	.87	.83
Rhythm Q178	.68	.59	.62	.79	.72
Contours Q17	.47	.41	.39	.49	.58
Contours Q19	.22	.21	.20	.16	.34
Contours Q20	.91	.92	.90	.96	.86
Composers Q30	.31	.32	.24	.37	.29
Composers Q32	.27	.16	.22	.32	.36
Composers Q38	.95	.94	.98	.94	.93
Styles Q41	.57	.45	.55	.73	.54
Styles Q43	.80	.82	.83	.76	.77
Styles Q44	.29	.15	.20	.39	.39

As a precursor to performing more in-depth analyses on the subtests, I created a correlation matrix of the **five subtest scores** (the five factors) [Table 5]. Since the MDAST was both an *aptitude **and** an achievement* assessment, knowing the values for each subtest was essential, as each subtest had to be capable of standing on its own.

Table 5 represents the findings on a "Subtest Correlation Matrix (n=357)." The findings sample ranged from 0.088 (Contour and Composer) to 0.405 (Composer and Style). Upon examination of the findings, there are only two samples above the .100s: 0.323 (Pitch and Rhythm) and 0.405 (Composer and Style). Both correlations are low, but the rest of the values manifest extremely low correlations between the subtests. This table is evidence that there are too many poor items in these subtests.

Table 5.

Subtest Correlation Matrix (n=357)

	Subtests				
Subtests	1. Pitch	2. Rhythm	3. Contour	4. Composer	5. Style
1. Pitch	1.000				
2. Rhythm	.323	1.000			
3. Contour	.144	0.194	1.000		
4. Composer	.195	0.170	0.088	1.000	
5. Style	.151	0.188	0.152	0.405	1.000

Note. All correlations were statistically significant ($p < .01$) except the correlation between contour and composer.

4.3 Research Question Results

My work was guided by a set of research questions focused on the procedures and analyses encapsulated within my study. They represented the order and content necessary to stay on track and provided my dissertation with a logical and sequential design.

4.3.1 *Research Question #1a*

#Q1a. What is the strength of the evidence supporting the validity of the *Musical Discrimination and Styles Task* (MDAST)—with content validity evidence provided by a panel of experts?

A panel of experts already reviewed the long version of this measure to determine adequate content validity. At that time, I replaced seven items in the long MDAST. During the item evaluation of the *MDAST **Long** Version*, I removed any item that did not have an 80% agreement on the correct answer from the test. Then, the item was replaced and reevaluated. However, experts did not review the short and abbreviated versions since they weren't created yet. Since I intended to use the *MDAST **Abbreviated** version* for the dissertation study this time, I needed a new panel of experts to reevaluate this newer version before the test could progress.

Therefore, the first logical step toward establishing content validity entwined it with achieving data to prepare the study for use. As was demonstrated in the pilot study, one accurate method of ensuring content validity in the MDAST was to submit the assessment for review by a **new panel of experts with terminal degrees**. The MDAST Long went through this process in the pilot study, albeit concisely. The **new experts** (n=5) took the measure through the same process expected for the student sample. During item evaluation on the MDAST, the researcher

removed any item that did not have an 80% agreement on the correct answer from the test. Then, the item was replaced and reevaluated.

For the dissertation study, a new panel of experts participated in a more thorough review of the current study [Table 6]. They still took the MDAST, but the Abbreviated Expert version allowed for more expert feedback on each question. When analyzing the MDAST, ***the items still had to reach an 80% agreement on the correct answer*** to be deemed acceptable. Each item still featured the audio samples necessary to complete it and the choices for consideration (same, different, or 'I don't know'). Additionally, beneath the student item, a second item appeared that featured a Likert scale with a question for further consideration: "Was the item before this one easily understood?" (1=Very easily, 2=Easily, 3=OK, 4=Somewhat difficult, 5=Very difficult). Individual MDAST student items were ***not*** analyzed utilizing the *Kuder-Richardson Formula 20* as was previously intended. *However, I determined the reliability of the instrument through expert* analysis. Instead, the Likert items were used alone as a separate analysis to determine the test items' status. ***Any item that received a mean of 4 (or a .80 for "Somewhat difficult") was cut from the final test*** and replaced with items that passed these criteria. The results of this process revealed an instrument with a ***high content validity based on the high reliability that the experts determined of the instrument.***

The results from [Table 6] showed that the MDAST had a high level of content validity. However, as a researcher, I was concerned that most of the tests contained 1s and 2s for the experts' opinions regarding ease of understanding. I had to ponder whether the MDAST was too easy for the students. Chapter 5 discusses this subject.

Table 6.

MDAST DISSERTATION: Content Validity by Panel of Experts

Item	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Q5	1	2	1	2	1
Q6	1	1	2	1	1
Q7	1	2	1	1	1
Q176	1	1	2	1	1
Q177	2	2	1	1	1
Q178	3	3	2	2	2
Q17	1	2	2	1	2
Q19	3	2	3	2	2
Q20	2	1	1	1	2
Q30	1	2	2	2	1
Q32	2	1	2	2	2
Q38	1	1	1	1	1
Q41	2	2	2	1	1
Q43	2	2	2	2	2
Q44	1	1	2	1	1

Note. "Was the item before this one easily understood?" (1=Very easily, 2=Easily, 3=OK, 4=Somewhat difficult, 5=Very difficult).

4.3.2 *Research Question #1b.*

#Q1b. What is the strength of the evidence supporting the validity of the *Musical Discrimination and Styles Task* (MDAST)—with internal structure validity evidence, as provided by confirmatory factor analysis?

Confirmatory factor analysis was used to evaluate the five-factor model underlying the MDAST. Analyses were conducted using Mplus. Because the 15 items were categorical (0=incorrect, 1=correct), weighted least squares with mean and variance adjustment estimation (WLSMV) was used to estimate the model. Results indicated that the model had poor fit, $\chi 2$ (80, N = 357) = 225.38, p <.001. Alternative measures of fit also indicated that the model had poor fit (RMSEA = .071, CFI = .722, SRNR = .145). Hair et al. (2014, p. 618) proposed that "standardized loading estimates should be .5 or higher, and ideally .7 or higher." Only nine items had loadings greater than .50, and only three were greater than .70. Due to confirmed poor fit through several alternate methods, and the amount of items with low standardized loadings, this CFA affirms that the MDAST Abbreviated is not ready for use in the classroom setting at this time. Future studies will be done to address these issues by attending to item diagnostics, replacements, and a new panel of experts.

Table 7.

Confirmatory Factor Analysis Results (n=357)

Items	Standardized Loading	Standard Error
Pitch Q5	.679	.083
Pitch Q6	.621	.081
Pitch Q7	.611	.094
Rhythm Q176	.451	.115
Rhythm Q177	.574	.109
Rhythm Q178	.604	.101
Contours 17	.101	.215
Contours 19	-.030	.080
Contours 20	.264	.554
Composers 30	.566	.070
Composers 32	.887	.073
Composers 38	.198	.128
Styles 41	.726	.067
Styles 43	-.342	.091
Styles 44	.862	.058

Examining the internal structure of the MDAST determined the degree to which the relationships between test items and test sections (or subtests) corresponded to the construct on which I based the proposed test score interpretations. For example, the MDAST's conceptual framework included several components that are supposed to be homogeneous but distinct from one another. The degree to which the item interrelationships supported the framework's assumptions is essential to establishing the MDAST's validity (American Educational Research Association et al., 2014).

The MDAST concluded a small pilot study with (n=7). However, the study results were inconclusive due to the lack of participant involvement. This current study had a much larger

sample to ensure a more valid and reliable result. The minimally projected (n=260) was to yield a power of .80, which was sufficient for the study to complete at least a CFA for this study. While I reached my power threshold of 260 participants with 362, it was not a sufficient sample to complete both an exploratory and confirmatory analysis at this time. Therefore, the exploratory factor analysis will have to wait until I undertake a future study with the MDAST after graduation. I chose to conduct a CFA with this study and must wait to undertake any other SEM with the MDAST in future research after graduation.

4.3.3 *Research Question #1c*

#Q1c. What is the strength of the evidence supporting the validity of the Musical Discrimination and Styles Task (MDAST)—with "relations to other variables" as provided by examining the relationships between **grade level** and the **subtests**?

Building on the CFA model, I added the *independent variable*, ***grade level*** (grades 3, 4, 5, and 6), as a predictor of the *dependent variable*, each of the five ***subtests*** (pitch, rhythm, contour, composers, and styles). Grade level had statistically significant positive relations with pitch (unstandardized coefficient = .132, standard error = .052, $p = .011$), composers (unstandardized coefficient = .094, standard error = .042, $p = .025$), and styles (unstandardized coefficient = .179, standard error = .047, $p < .001$). Students in higher grades scored higher on each of these three subtests. These statistically significant positive relations with three items in the dependent variable, subtests (pitch, composers, and styles), suggest that validity in the form of "relations to other variables" exists between them and the independent variable, grade level, especially for students in higher grades.

4.3.4 *Research Question #2a*

Since the students chose their answers on a same/different/"I don't know" basis, I used the Kuder-Richardson 20 formula to analyze for reliability. "I don't know" is counted as a wrong answer in this analysis to allow for the dichotomous relationship of the KR-20 (The "I don't know" response is also to prevent students from guessing).

The data analysis step occurred after the participants took the measure, and I determined the instrument's reliability. I utilized the KR-20 to analyze the **student version** of the test (without any Likert items).

Individual subtests had their own KR-20 analyses run to determine their reliability. It was crucial in establishing the overall reliability of the MDAST that I evaluated each section independently. The individual **subtest** KR-20s and the overall KR-20 represented a comprehensive look at the MDAST's **true reliability** [Table 8]. Additionally, the subtests underwent bivariate correlation (Pearson's) to determine the strength of their relationship to one another [Table 5].

Table 8.

Reliability Statistics (n= 357)

Grade Level 3 - 6	Cronbach's Alpha	N of Items
0. Total	.681	15
1. Pitch	.449	3
2. Rhythm	.398	3
3. Contours	.118	3
4. Composers	.346	3
5. Styles	.056	3

Note. Subtest order: 1- pitch discrimination, 2- rhythm discrimination, 3- melodic contours, 4- musical composer comparison, 5- musical style comparison

4.4 Summary

This chapter was the culmination of the analyses done after collecting data in the form of student scores. First, I described the data set, how it manifests in this study (item difficulty and subtest correlation matrix), and the data's demographics. Then, I provided the analysis results for each research question conducted in the study. In the final chapter, I will interpret these results, draw conclusions from them, and make recommendations about how the knowledge from this study might springboard new works in both current practice and future research.

Chapter 5: Summary, Interpretations and Recommendations

In this final chapter, I discussed how the research, studies, analyses, and theoretical framework together into a cohesive whole. First, in the (5.1) Summary section, I described the two purposes I used to keep the dissertation focused on the right track, and I gave a quick reminder of the MDAST's connection to Aptitude and Achievement. Second, in the (5.2) Interpretations section, I explained the Fundamental Knowledge gained from the study, why it may be necessary for Music Education, and how the Reliability and Validity scores of the MDAST apply to research or the classroom. Next, in the (5.3) Recommendations section, I made suggestions regarding Implications for Practice and Implications for Future Research. In conclusion, I offered my Final Thoughts on this study and its meanings to me.

5.1 Summary

This dissertation began as a classroom assignment to create a 10-item assessment in a Measurement class with Dr. C. Victor Fung. As a part of that assignment, we had to write about what we created, and we needed to demonstrate one form of analysis each for validity and reliability. As soon as I began making the items for the test, I made a couple of rules for myself to help me stay on track. The **first** purpose was to document the development of a new music aptitude and achievement instrument entitled the *Musical Discrimination and Styles Task* (MDAST), and the **second** purpose was to describe the strength of evidence supporting the validity and reliability of this developmentally appropriate music achievement instrument for elementary and middle school music students.

I thought I was creating an aptitude test for the music classroom. I filled out all the requirements Dr. Fung gave us to get an "A" on the assignment, and I just kept building on it. Before I knew it, I had an assessment of sixty items I had either composed myself (I recorded and uploaded) or collected as 10- to 15-second clips for listening samples. I built the entire test on the Qualtrics platform within the semester and wanted to try it out to see if it worked in real-life applications. I did a test for *content validity* (which fulfilled the validity requirement) by having the Measurement class take the test (since we are all students for terminal degrees), and ran the KR-20 to test for *reliability*, which gave me my first data before the end of the semester.

5.1.1 The Two Purposes

Even from the early stages in the MDAST development, I followed the **first** purpose that I described with this dissertation. I documented everything I did with a "study diary" in which I notated all the actions I took with the MDAST. I detailed how I created the items. I expressed what I thought worked well and what needed to go back for revisions. I kept ideas I'd gotten from professors, books, journals, websites, colleagues, and more. Then, I wrote, wrote, and wrote in this document about everything pertaining to my study, about everything I had learned from my pilot study, my research for the literature review, and every step I took as I worked on this massive project that has spanned the last few years of my doctoral journey. As I finished documenting the end of this dissertation, the first purpose was fulfilled.

The **second** purpose of this dissertation was to describe the strength of evidence supporting the validity and reliability of the MDAST. After reviewing the results of the analyses performed in "Chapter 4," I must conclude that the MDAST is not ready for classroom use at this time. First, evaluation of the "Item Difficulties" revealed that there were many flawed items that

needed revision. Some of the items in question scored low, not because they were too hard, but because they were too easy. That was something I had heard of in theory, but this was the first time I had experienced it for myself. Next, the CFA confirmed poor fit through several alternate methods, and the number of items with low standardized loadings was poor, as well. These flawed items would need to be replaced. However, when I used the CFA to determine if there was a relationship between "grade level" and "subtests," I found positive results—there were statistically significant positive relations with three items in the dependent variable, subtests (pitch, composers, and styles), suggesting that validity in the form of "relations to other variables" exists between them and the independent variable, grade level, especially for students in higher grades. This result suggests that as students grow up by grade level, their success level is positive for three of the subtests: pitch, composers, and styles.

5.1.2 Aptitude and Achievement

After researching these subjects at length, I have not found another music assessment that contains both aptitude and achievement tests grouped together as one unit. Therefore, the MDAST is something new to music assessments. I believe the MDAST can fill a need in the music classroom that does not currently exist—an aptitude and achievement test lasting a combined total of 20 minutes for music educators to track student abilities and progress from year to year.

5.2 Interpretations

Interpretations are ways of explaining data or phenomena. This section of the dissertation dealt with explaining two essential discussion subjects. The first one dealt with the

"*Fundamental Knowledge From the Study*," and the second subject area to discuss and explain was how the data transformed into results that established the "*Reliability and Validity of the MDAST*. These two discussion areas stemmed from the knowledge gained after conducting the study and the results obtained after running the analysis connected to the research questions.

5.2.1 Discussion: Fundamental Knowledge From the Study

When I began working on this discussion section, I was unsure how to extract a subject this broad. Therefore, I toured my dissertation and tried to ask pertinent questions about what I had learned while working on this incredible project. The short discussion that follows is my attempt to answer those questions.

After comparing and contrasting my findings with other studies on similar subjects, I evaluated how my research related to them. I was frustrated to reaffirm that few music researchers are creating new music achievement assessments. I saw some similarities and differences when I compared my study to the studies of the three researchers I discussed in my paper. I could also see that according to my reliability results [Table 8], I had years of work to do to reach their level, and many more improvements to make and future studies to do before the MDAST can be a normed, referenced test.

I believe I had different results because I have not made corrections yet to remove faulty items and replace them with improvements, run subsequent studies, or put in the same level of time as those researchers did. The creation of a new assessment like this one takes years to perfect. That said, I was encouraged to see the test improved over the pilot version, which leads me to believe I am on the right track. As a result of this study, I believe the CFA findings regarding the validity of the three "subtests" by "grade level" were encouraging. I think the

knowledge gleaned from this study is the first step in a line of research that is possible to continue upgrading after graduation and beyond.

5.2.2 *Discussion: Reliability and Validity of the Musical Discrimination and Styles Task*

Before I ran any analyses for reliability or validity, I performed an evaluation of the "Item Difficulties" to determine the individual strength of each item. This analysis revealed many flawed items needing revision. After review of individual items, I noted that some of the items in question scored low, not because they were too hard, but because they were too easy.

5.2.2.1 Reliability of the MDAST

I determined the reliability of this study by running the student data through the Kuder-Richardson 20 formula (a variation of Cronbach's Alpha) on both the comprehensive (whole) study level (including all subtests together) and the individual subtest level. I also conducted a subtest analysis. By testing the individual **subtest** KR-20s independently and including the **overall** KR-20 in the results display, the research represented a comprehensive look at the MDAST's **true reliability**. [Table 8]

When interpreting Cronbach's Alpha scores*(reliability*), George and Mallery (2003), are cited often and provide these scores as a guide for interpretation: *α > 0.9 (Excellent), > 0.8 (Good), > 0.7 (Acceptable), > 0.6 (Questionable), > 0.5 (Poor), and < 0.5 (Unacceptable).* The highest score for the *MDAST Abbreviated* was for the **Total** score (.681), which, according to George and Mallery (2003), would be "Questionable." Every subtest score was below 0.5 (Unacceptable).

The MDAST's reliability score is deceiving if one only looks at the overall score because it shows that the overall test is still high enough to be in the **Questionable** category. It is considerably higher than the subtest scores (**Unacceptable**), which shows where the real work needs to occur. These scores look discouraging; however, this will be a salvageable study with the correct modifications. Some deep flaws need correction on the fundamental level—the test items. They need replacing, so I will turn to the Item Difficulties Statistics and coordinate the faulty items with problem items on the CFA and Cronbach's Alpha to determine which items I must remove to fix the errors (on both the total and subtest levels). Then, I will create a short "test" of items that would be eligible to replace the faulty items with new items from my test bank. I will select a new Panel of Experts to evaluate the short "eligibility test" with the possible replacement items and determine which new items are acceptable to add to the real test (Any leftover items can serve as future replacement items if the need arises.). After I add those new items, I will have the Panel of Experts take the new and improved MDAST Abbreviated. I will repeat this process as often as necessary until all the test items pass the Panel of Experts procedure and the MDAST is approved. This process will bring up Cronbach's Alpha and raise the reliability of the study. The MDAST will be prepared for the subsequent research study throughout this process.

5.2.2.2 Validity of the MDAST

First, evaluating the "Item Difficulties" revealed many flawed items that needed revision. Some of the items in question scored low, not because they were too hard but because they were too easy. I had heard of this in theory, but this was the first time I had experienced it for myself.

Next, the CFA confirmed poor fit through several alternate methods, and the number of items with low standardized loadings was poor, as well. These flawed items would need to be

replaced. However, when I used the CFA to determine if there was a relationship between "grade level" and "subtests," I found positive results—there were statistically significant positive relations with three items in the dependent variable, subtests (pitch, composers, and styles), suggesting that validity in the form of "relations to other variables" exists between them and the independent variable, grade level, especially for students in higher grades. This result suggests that as students grow up by grade level, their success level is positive for three of the subtests: pitch, composers, and styles.

5.3 Recommendations

"Recommendations" (within the research paradigm) are possible suggestions that consider the study results, the literature, and limitations that may bind either of them. "Implications" are slightly different, however. While they fall under the heading of "Recommendations" because they meet the broad definition, they are more specific. They can represent possible options for existing research modifications or suggest new ideas in practice, theory, and policy. Implications may also present new initiatives for future research inspired by the current study, additional problems discovered, or fresh research questions raised by the current results. As the MDAST study progressed, I kept a list of ideas that met these criteria. Therefore, the sections below are my recommendations for the implications that I felt were applicable to this study.

5.3.1 *Implications for Music Practice*

The idea I am suggesting here was an absolute surprise despite being right in front of me for my entire doctoral student career. I struggled with how I would be able to go from a research

form of the test that required anonymity to a form that I'd be able to keep the same paradigm but add on an individual student identity component for grading accountability and annual tracking. I had an epiphany when I realized I had been tracking my sons' scores on the MDAST annually on scantron sheets.

It is possible for an entire school to allow its students to take the assessment online. While students completed the online portion of the test, they would bubble the answers they chose online onto the scantron sheet on their desks. Meanwhile, the research portion of the MDAST would continue online without any protocol being breached. The teacher would collect the scantrons, which are machine-gradable. I could provide them with a program to keep track of the student scores. By keeping records of the scores, it would be easy to track students' music achievement and growth rate. I could eventually envision many more schools willing to participate in a study that could provide them with that amount of data and results.

Once a music educator has the students' achievement scores, they can use them to help teach their students on a more differentiated basis. For instance, if a student consistently scores low on their achievement test, they may have difficulties in music and need extra help. If a student consistently scores high on their achievement test, you could use that student as a teaching assistant or as a peer buddy for the student who is struggling (pair them up in class).

5.3.2 *Implications for Future Research*

A study is never broad enough to encompass all the ideas a researcher wants to accomplish within it. Inevitably, some items must be added to a "To Do" list for Future Research. Still, having this list allows a researcher to continue to process the next great thing they can work on and eventually bring to fruition. When a researcher continues to accomplish

enough items on the list, they become known for that subject matter, and when they only do research from that list, it becomes their life's work.

One research study I would like to accomplish soon is translating the MDAST into Spanish. I firmly believe in diversity and am always upset when I see how difficult it is for English Language Learners to take the required tests every year. Once I get the reliability and validity to respectable levels, this is the next thing I want to tackle.

In another research idea, I'd like to do a mixed methods study that maintains only enough quantitative research to preserve essential reliability and validity while making the qualitative techniques the primary focus of the research. The study would continue to access and analyze the survey questions at the end of the MDAST for basic and immediate feedback from students as they finish testing. However, the primary data collected would be recordings of individual face-to-face student interviews with students who took the test. The video recordings would be the primary data source for future analysis with DeDoose. Then, I would analyze the video and audio transcripts for codes and emergent themes as described in cognitive interviewing techniques (Miller, 2011).

Another possible research idea is based on the scores by Grade Level. One observation of note is that the fifth-grade scores suddenly dropped downward, even below the third-grade. This sudden drop made me curious about the cause, and I would need to do more research to find out why it occurred. However, I wondered if a partial cause might be the onset of puberty, causing the body's growth and the sudden expansion of the human brain (causing cognitive processing to go through significant changes). This study may require extensive literature research to discover if there are similarly documented occurrences among other assessments of note. Then, the MDAST may need to be repeated several times in different settings to determine if the original

fifth-grade observation was an aberration or a legitimate change in cognitive processing and performance.

In another study idea, the literature suggests that students in Grades 1 – 5 experience a faster level of growth between grades than students in Grades 6 – 8, and that by age nine, our music aptitude reaches its peak (Gordon, 1970). I am interested in a study to see if the faster level of growth data between the younger students is provable, and if it is, to see if any musical phenomena occur around age nine that might corroborate the difference between when the two Grade Level groups' speed level spreads apart.

5.4 Final Thoughts

This dissertation had two purposes that helped me focus and remain on track. The ***first*** purpose was to ***document the development*** of a new music achievement instrument, the MDAST (Colwell, 1970), and that purpose succeeded. The ***second*** purpose was to describe the ***strength of evidence*** supporting the ***validity and reliability*** of the study (Gordon, 1970). After conducting appropriate analyses of reliability and validity, the results showed that the reliability was *poor* and *unacceptable*, and the test has an overall status of "poor fit." The reliability and validity showed flaws in the test items that need replacement. However, the validity (through CFA) also revealed the "relations between other variables" as confirmed between the independent variable, "grade level" (3, 4, 5, 6), and the three of the five dependent variables, "subtests" (pitch, composers, and styles).

At the beginning of this journey, I was hooked on the idea of making the MDAST a real test and eventually having it be usable in a real-world classroom. However, I never preplanned that this test would be my dissertation. Nevertheless, the MDAST and the studies that followed it

became the years of work that defined my doctoral degree. There have been times when I completely immersed myself in the research and loved it. There have also been days when I wondered why I ever chose this subject. But since I reached the end of my dissertation research, I was excited to see the fruit of my work. When I started this project, I wanted to create a test that could eventually be used in the classroom, and now I feel like I am much closer to reaching that goal.

References

Addessi, A. R., Baroni, M., Luzzi, C., & Tafuri, J. (1995). The development of musical stylistic competence in children. *Bulletin of the Council for Research in Music Education*, *127*, 8–15. https://www.jstor.org/stable/40318759

Alex, D. (2023). What are Implications and Recommendations in Research? How to Write It, with Examples. *Semantic Scholar Online Database.* Retrieved June 7, 2024, from http://saber2.net/Archivos/Schema-Theory-Intro.pdf

Algarabel, S., & Dasi, C. (2001). The definition of achievement and the construction of tests for its measurement: A review of the main trends. *Psicologica*, *22*(1), 43–66. https://www.redalyc.org/pdf/169/16917746001.pdf

Algassen, M., Kliegel, M., & Williams, T. (2005). Pitch perception in children with autistic spectrum disorders. *British Journal of Developmental Psychology*, *23*, 543–558. Retrieved March 22, 2020, from https://onlinelibrary.wiley.com/doi/full/10.1348/026151005X26840?casa_token=GOPLfUt62AsAAAAA%3A3u2IUTekPlBuzj0EyhcJoaCJF7AxcyzINGMPQymBXhFcSKEKIERp09urfnIJDI3YFxtR1cTPjZciB7ti

Aliferis, J. (1957a). The Aliferis Music Achievement Test. *American Music Teacher*, *7*(1), 6–20. Retrieved December 12, 2023, from http://www.jstor.org/stable/43536128

Aliferis, J. (1957b). The Aliferis Music Achievement Test. *American Music Teacher*, *7*(1), 6–20. Retrieved September 28, 2022, from http://www.jstor.org/stable/43536128

Aliferis, J., & Kraehenbuehl, D. (1958). Re: The aliferis music achievement tests. *Journal of Music Theory*, *2*(2), 237. https://doi.org/10.2307/843200

American Educational Research Association, Association, A. P., & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Andrews, M. L., & Madeira, S. S. (1977). The assessment of pitch discrimination ability in young children. *Journal of Speech and Hearing Disorders*, *42*(2), 279–286. https://doi.org/10.1044/jshd.4202.279

Applied Doctoral Center, Section 1, Problem Statement. (2024). Retrieved June 25, 2024, from https://resources.nu.edu/c.php?g=1013602&p=7638573

Bruin, J. (2006). Newtest: Command to compute new test. UCLA: Statistical Consulting Group. https://stats.oarc.ucla.edu/stata/ado/analysis/

Brophy, T. S. (2019). Assessment in music education: The state of the art. In T. S. Brophy (Ed.), *The oxford handbook of assessment policy and practice in music education, volume 2* (pp. 901–932). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190248130.001.0001

Cangro, R. M. (2014). Measuring music education: Music education assessment in Illinois. *Music Educators Journal*, *100*(3), 20–21. Retrieved September 8, 2022, from https://doi.org/10.1177/0027432113519929

Colwell, R. (1970). The development of the music achievement test series. *Bulletin of the Council for Research in Music Education*, 57–73. https://www-jstor-org.ezproxy.lib.usf.edu/stable/40317115?seq=1

Colwell, R. (2019a). An overview of music tests and their uses. In T. S. Brophy (Ed.), *The Oxford handbook of assessment policy and practice in music education, Volume 1* (pp. 536–573). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190248093.013.24

Colwell, R. (2019b). Assessment policy in music education. In T. S. Brophy (Ed.), *The Oxford handbook of assessment policy and practice in music education, Volume 1* (pp. 2–32). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190248093.013.1

Colwell, R. J. (1969). *Manual: Music achievement tests 1 - 4*. Follet Publis Company Educational Corporation.

Comeau, G. (2009*). Piano pedagogy: A research and information guide*. New York: Routledge.

Conway, C. M. (2020). *Teaching music in higher education* (2nd ed.) [e-book]. Oxford University Press. https://doi.org/10.1093/oso/9780190945305.003.0002

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

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory* (1st ed.). Holt, Rinehart, and Winston.

Dang, S. (2020). Re: What to do with cases of cross-loading on Factor Analysis?. Retrieved from: https://www.researchgate.net/post/What-do-do-with-cases-of-cross-loading-on-Factor-Analysis/5efd7bbd6b9aab36ac600a0c/citation/download

Dattalo, P. (2013). Multivariate multiple regression. In *Analysis of multiple dependent variables* (pp. 87–108). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199773596.003.0004

de Manzano, Ö., & Ullén, F. (2021). Domain specific traits predict achievement in music and multipotentiality. *Intelligence*, *89*, 101584. https://doi.org/10.1016/j.intell.2021.101584

Dewey, J. (1910). *How we think*. D C Heath. https://doi.org/10.1037/10903-000

Duell, O. K., & Anderson, R. C. (1967). Pitch discrimination among primary school children. *Journal of Educational Psychology*, *58*(6, Pt.1), 315–318. https://doi.org/10.1037/h0025223

Erkut, S. (2010). Developing multiple language versions of instruments for intercultural research. *Child Development Perspectives*, *4*(1), 19–24. Retrieved February 17, 2024, from https://doi.org/10.1111/j.1750-8606.2009.00111.x

Erlings, B. (1977). A pilot investigation of relationships between elementary keyboard sightreading achievement by music majors in college and selected musical profile tests. *Bulletin of the Council for Research in Music Education*, *50*(Spring), 14–17. Retrieved December 12, 2023, from https://www.jstor.org/stable/40317438

Fancourt, A., Dick, F., & Stewart, L. (2013). Pitch-change detection and pitch-direction discrimination in children. *Psychomusicology: Music, Mind, and Brain*, *23*(2), 73–81. https://doi.org/10.1037/a0033301

Fyk, J. (1995). Musical determinants of melodic contour recognition: Evidence from experimental studies of preschoolers. *Bulletin of the Council for Research in Music Education*, *127*, 72–79. https://www.jstor.org/stable/40318769

Gallagher, F. D. (1971). *A study of the relationships between the Gordon music aptitude profile, the Colwell music achievement tests, and the Indiana - Oregon music discrimination* [e-book]. Indiana University. https://www.proquest.com/docview/302612749?pq-origsite=gscholar&fromopenview=true

Gardner, H. (1973). Children's sensitivity to musical styles. *Merrill-Palmer Quarterly of Behavior and Development*, *19*(1), 67–77. https://www.jstor.org/stable/23083794

George, D. & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. 11.0 update (4th ed.). Allyn & Bacon. https://www.proquest.com/docview/199735362?accountid=14745&parentSessionId=j18my%2BuJLRH8ACEKVZEbZdDRA%2Bl5BaiL8WhA0rzsyb4%3D&pq-origsite=primo&sourcetype=Trade%20Journals

Gordon, E. E. (1965). The musical aptitude profile: A new and unique musical aptitude test battery. *Bulletin of the Council for Research in Music Education*, *6*, 12–16. https://www.jstor.org/stable/40316898

Gordon, E. E. (1970a, 1991). *Iowa Tests of Music Literacy*. Chicago: GIA.

Gordon, E. E. (1979). *Manual for the Primary Measures of Music Audiation*. Chicago: GIA.

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

Gordon, E. E. (1989). *Manual for the Advanced Measures of Music Audiation*. Chicago: GIA.

Gordon, E. E. (1995b, 1988, 1965). *Manual: Musical Aptitude Profile*. Chicago: GIA.

Gordon, E. E. (1999). All about audiation and music aptitudes. *Music Educators Journal*, *86*(2), 41–44. https://doi.org/10.2307/3399589

Gordon, E. E. (2011). *Roots of Music Learning Theory and Audiation* [e-book]. University of South Carolina Scholar Commons. https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1000&context=gordon_articles

Gordon Institute for Music Learning & Gordon Institute for Music Learning. (2017). *AUDEA: A Journal for Research and Applications of Music Learning Theory, 19(1)* [e-journal]. GIML. https://giml.org/wp-content/uploads/Audea_Vol19No1_Spring2017.pdf

Hargreaves, D. (2012). The development of artistic and musical competence. In *Musical beginnings, origins, and development of musical competence* (pp. 144–170). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198523321.003.0006

Hargreaves, D. J., & Galton, M. (1992). Aesthetic Learning: Psychological theory and educational practice. In B. Reimer & R. A. Smith (Eds.), *National Society for the Study of Education yearbook on the arts in education* (pp. 124–150). NSSE. https://doi-org.ezproxy.lib.usf.edu/10.1037/h0094169

Hargreaves, D., & Lamont, A. (2017). *The psychology of musical development* (1st ed.). Cambridge University Press.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve?. *International journal of human-computer interaction*, *32*(1), 51-62. [PDF] tandfonline.com

Huitt, W., & Hummel, J. (2003). Piaget's theory of cognitive development. *Educational Psychology Interactive*. https://intranet.newriver.edu/images/stories/library/stennett_psychology_articles/Piagets%20Theory%20of%20Cognitive%20Development.pdf

Ilari, B., & Polka, I. (2006). Music cognition in early infancy: Infants' preferences and long-term memory for Ravel. *International Journal of Music Education*, *24*, 7–20. https://doi.org/10.1177/0255761406063100

Jensen, M. C. (2016). *Measuring music reading: A guide to assessment methods* [Master's thesis, University of Ottowa]. Universite d'Ottawa/University of Ottawa. https://ruor.uottawa.ca/items/5c2c55f0-46d2-4a57-9f77-1d6722173b30

Kaur, G., Anand, D., & Awasthy, S. (2016). Intelligence and aptitude testing. In *Military psychology: Concepts, trends and interventions* (pp. 33–55). SAGE Publications Pvt. Ltd. https://doi.org/10.4135/9789353885854.n3

Kornicke, L. E. (1992). *An exploratory study of individual difference variables in piano sight-reading achievement* (Order No. 9301458). Available from ProQuest Dissertations & Theses Global. (304007507). Retrieved from http://search.proquest.com/docview/304007507?accountid=14701

Kraehenbuehl, D. (1957). Re: The aliferis music achievement tests. *Journal of Music Theory*, *1*(2), 208. https://doi.org/10.2307/843280

Law, L. N. C. & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PloS one, 7*(12), 1-15.

Levy, K. (2001). *Music readers and notation: Investigation of an interactive model of rhythm reading* (UMI No. 3034120).

MacKnight, C. B. (1975). Music reading ability of beginning wind instrumentalists after melodic instruction. *Journal of Research in Music Education, 23*(1), 23-34.

Marshall, N. A., & Hargreaves, D. J. (2007). Musical style discrimination in the early years. *Journal of Early Childhood Research*, *5*(1), 32–46. https://doi.org/10.1177/1476718x07072150

May, W. V. (1985). Musical style preferences and aural discrimination skills of primary grade school children. *Journal of Research in Music Education*, *33*(1), 7–22. https://doi.org/10.2307/3344754

McLeish, J. (1968). The validity and reliability of bentley's measures of musical abilities. *British Journal of Educational Psychology*, *38*(2), 201–201. Retrieved September 10, 2022, from https://doi.org/10.1111/j.2044-8279.1968.tb02007.x

Miller, R. E. (1988). *Contributions of selected music skills to music sight-reading achievement and rehearsed reading achievement* (Doctoral dissertation, University of Illinois at Urbana-Champaign, 1988). (UMI No. 8908779)

Musumeci, O. (1997). The Bentley "Measures of Musical Abilities: as predictors of conservatory students' musical achievement. *Bulletin of the Council for Research in Music Education*, *133*, 95–99. Retrieved September 8, 2022, from http://www.jstor.org/stable/40318846

Miller, K. (2011). Cognitive interviewing. *Question evaluation methods: Contributing to the science of data quality*, 49-75.

Newman, I., Ridenour, C. S., Newman, C., & DeMarco, G. M. P. (2003). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 167–188). Sage.

Palmer, M. (1976). Relative effectiveness of two approaches to rhythm reading for fourth-grade students. *Journal of Research in Music Education, 24*(3), 110-118.

Petzold, R. G. (1963). The development of auditory perception of musical sounds by children in the first six grades. *Journal of Research in Music Education*, *11*(1), 21–54. https://doi.org/10.2307/3344529

Piaget, J. (1969). *The psychology of the child* (lst ed., later printing). Routledge & Kegan Paul PLC.

Pick, A. D., Palmer, C. F., Hennessy, B. L., Unze, M. G., Jones, R. K., & Richardson, R. (1988). Children's perception of certain musical properties: Scale and contour. *Journal of Experimental Child Psychology*, *45*(1), 28–51. https://doi.org/10.1016/0022-0965(88)90049-5

Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, *26*(1), 108–116. http://www.psicothema.com/pdf/4168.pdf

Rowntree, J. P. (1970). The Bentley "Measures of Musical Abilities" – A critical evaluation. *Bulletin of the Council for Research in Music Education, 22*(1), 25-32.

Schellenberg, E., & W. Weiss, M. (2012). Music and cognitive abilities. In *The psychology of music* (pp. 499–550). Elsevier. https://doi.org/10.1016/b978-0-12-381460-9.00012-2

Seashore, C. E. (1915). *The Measurement of Musical Talent* [e-book]. G. Schirmer. https://www.google.com/books/edition/The_Measurement_of_Musical_Talent/qL-5AAAAIAAJ?hl=en&gbpv=1&printsec=frontcover

Shuter-Dyson, R., & Gabriel, C. (1981). *Psychology of musical ability* (2nd ed.). Methuen.

Snow, R. E. (1991). The concept of aptitude. In *Improving inquiry in the social sciences* (pp. 249–284). https://books-google-com.ezproxy.lib.usf.edu/books?hl=en&lr=&id=AZyf_xDrzrUC&oi=fnd&pg=PA249&dq=what+is+aptitude&ots=rwhrGC3w49&sig=janFRyJumq5JYTNuhgfi5e4idbg#v=onepage&q=what%20is%20aptitude&f=false

Sree, D. (2020). Musicality and musical sensitivity in children. *International Journal of Psychosocial Rehabilitation*, *24*(4), 6781–6789. https://doi.org/10.37200/ijpr/v24i4/pr2020489

Stanutz, S., Wapnick, J., & Burack, J. A. (2012). Pitch discrimination and melodic memory in children with autism spectrum disorders. *Autism*, *18*(2), 137–147. https://doi.org/10.1177/1362361312462905

Trochim, W. M. (2006). *Research Methods Knowledge Base* (2nd ed.) [e-book]. http://www.socialresearchmethods.net/kb/

Van Zee, N. (1976). Responses of kindergarten children to musical stimuli and terminology. *Journal of Research in Music Education*, *24*(1), 14–21. https://doi.org/10.2307/3345062

Wadsworth, B. J. (1996). *Piaget's theory of cognitive and affective development: Foundations of constructivism* (5th ed.). Longman Publishing.

Walker, R. (1987). Some Differences between Pitch Perception and Basic Auditory Discrimination in Children of Different Cultural and Musical Backgrounds. *Bulletin of the Council for Research in Music Education*, *91*, 166–170. https://www.jstor.org/stable/40318079

Websters, P. (2000). Broadening the concept of music aptitude: New approaches to theory and practice. *In Sixth International Conference on Music Perception and Cognition, Keele, UK, August 6-10*. https://www.escom.org/proceedings/ICMPC2000/poster1/Webster.htm

Wesolowski, B., & Wind, S. (2019). Validity, reliability, and fairness in music testing. In T. S. Brophy (Ed.), *The Oxford handbook of assessment policy and practice in music education, Volume 1* (pp. 437–460). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190248093.013.20

Widmayer, S. A. (2004). *Schema theory: An introduction*. Retrieved June 7, 2024, from https://www.semanticscholar.org/paper/Schema-Theory-%3A-An-Introduction-Widmayer/47b15487db915f62aec7a1f57c6f64c0c1c5234f

Young, W. T. (1973). The Bentley "Measures of Musical Abilities": A congruent validity report. *Journal of Research in Music Education*, *21*(1), 74–79. Retrieved September 9, 2022, from https://doi.org/10.2307/3343982

Young, W. T. (1976). A longitudinal comparison of four music achievement and music aptitude tests. *Journal of Research in Music Education*, *24*(3), 97–109. https://doi.org/10.2307/3345153

Zimmerman, M. P. (2007). Musical characteristics of children. *Visions of Research in Education*, *17*(1), 8. https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=1099&context=vrme

Appendix A: Institutional Review Board Approval

UNIVERSITY of SOUTH FLORIDA

APPROVAL

October 12, 2020

Dawn Mitchell White
6802 Hanley Rd
Tampa, FL 33634

Dear Ms. Dawn Mitchell White:

On 10/12/2020, the IRB reviewed and approved the following protocol:

Application Type:	Initial Study
IRB ID:	STUDY001566
Review Type:	Expedited 7
Title:	Musical Discrimination and Styles Task: Assessing the developmental phases of elementary-aged children's music perceptions
Funding:	None
IND, IDE, or HDE:	None
Approved Protocol and Consent(s)/Assent(s):	• Protocol Version 1 10-11-2020.docx; • Verbal Assent Version 1 10-06-2020.pdf; • Verbal Parental Permission Version 1 10-06-2020.pdf; Approved study documents can be found under the 'Documents' tab in the main study workspace. Use the stamped consent found under the 'Last Finalized' column under the 'Documents' tab.

Within 30 days of the anniversary date of study approval, confirm your research is ongoing by clicking Confirm Ongoing Research in BullsIRB, or if your research is complete, submit a study closure request in BullsIRB by clicking Create Modification/CR.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Your study qualifies for a waiver of the requirements for the documentation of informed consent for remote consenting as outlined in the federal regulations at 45 CFR 46.117(c).

UNIVERSITY of SOUTH FLORIDA

This research involving children as participants was approved under 45 CFR 46.404: Research not involving greater than minimal risk to children is presented.

Requirements for Assent and/or Permission by Parents or Guardians: 45 CFR 46.408 Permission of one parent is sufficient.

Assent will be obtained as outlined in the IRB application.

Sincerely,

Various Menzel
IRB Research Compliance Administrator

Appendix B: MDAST Long Version Auditory Item Listing

Item #	Discrim-ination	Item #	Rhythms	Item #	Contours	Item #	Composers	Item #	Styles
1	e4 – f4 DIFF	13	DIFF	26	SAME	38	*Scheherazade 1 *Devil Went Down to Georgia DIFF	50	*Bach- Toccata & Fugue in D Minor *Bach- Goldberg Variations SAME
2	a3 – a3 SAME	14	SAME	27	DIFF	39	*Mozart- Clarinet Concerto 1 *Weber- Clarinet Concerto 1 SAME	51	*Mozart- Symphony 29 Mvmt 1 *Hindemith- Symphony in Bb DIFF
3	d4 - c#4 DIFF	16	SAME	28	SAME	40	*In the Mood *Sing, Sing, Sing DIFF	52	*Le Tombeau de Couperin- Prelude *Mozart- Gran Partita Mvt 3 DIFF
4	f3 – g3 DIFF	17	SAME	29	DIFF	41	*Chant of the Mystics *Hymn of the Cherubim DIFF	53	*Japanese Flute *Traveler's Song SAME
5	b3 –b3 SAME	18	DIFF	30	DIFF	42	*Adagio for Strings *Agnus Dei SAME	54	*When the Saints Go Marching In *Charleston SAME
6	f#4 - g4 DIFF	19	SAME	31	DIFF	43	*My Favorite Things *Take Five DIFF	55	*Scheherazade 2 *Vivaldi – Spring DIFF

7	e3 – e3 SAME	20	DIFF	32	SAME	44	*Rite of Spring 2 *Firebird 2 SAME	56	*Yardbird Suite *Giant Steps SAME
8	bb3-c4 DIFF	21	SAME	33	SAME	45	*Guitar Concerto de Aranjuez *Vivaldi Lute Concerto in D Major DIFF	57	*Eruption *Capricho Arabe DIFF
9	c5 – c5 SAME	22	DIFF	34	DIFF	46	*Rhapsody in Blue 1 *Rhapsody in Blue 2 SAME	58	*Rain in Jiang Nan *Passacaglia for Violin & Viola DIFF
10	g3-ab3 DIFF	23	DIFF	35	SAME	47	*Firebird 1 *Prelude to the Afternoon of a Faun DIFF	59	*Bach-Goldberg Variations-Aria *Minuet in G SAME
11	f2 - eb2 DIFF	24	DIFF	36	DIFF	48	*Dvorak Cello Concerto 1 *Dvorak Cello Concerto 2 SAME	60	*Mozart-Concerto for Flute & Harp Mvmt 2 *Native Americans in Talinn DIFF
12	d4 –d4 SAME	25	SAME	37	DIFF	49	*Pavane for a Dead Princess *Rite of Spring 1 DIFF	61	*Beethoven-Symphony 6 Mvmt 1 *Debussey-La Mer DIFF

Note. Items are labeled as SAME or DIFF (different) to function as the answer key. Items 13 – 25 are composed items on the musical staff. Therefore, the answer key is provided. *New Note*: Q176, Q177, Q178, Q17, Q32, Q38, Q41, Q43, and Q44 were items included in a six-part Dataset for the MDAST.

Appendix C: Written/Vocal Instructions on the MDAST Long Version

Items used in the MDAST Abbreviated Version are highlighted below.

Subtest	Item #	Written Instructions
1 – Discrimination	Q1	Let's practice a couple of examples for the first group of questions. You will hear two sounds. Choose the word that explains what you hear. Are the notes the same or different? If you can't tell which answer is right, just choose I don't know.
	Q2	This example is the same kind as the last one. Listen to the next set of sounds and choose your best answer. Are these notes the same or different?
	Q3	For the rest of this section, your answers will count. The example you're going to hear is the same kind as the last two you tried. Listen to the next set of sounds and choose your best answer. Are these notes the same or different?
Q5, Q6, & Q7	Q4 – Q12	This example is the same kind as the last one. Listen to the next set of sounds and choose your best answer. Are the notes the same or different?
2 – Rhythms	Q13	Let's practice a couple of examples for the next group of questions. Play each arrow button. You will hear two sets of musical sounds. Choose the word that explains what you hear. Are the musical notes the same beat or different? If you can't tell which answer is right, just choose I don't know.
	Q14	This example is the same kind as the last one. Listen to the next two sets of sounds and choose your best answer. Are the sets of musical notes the same beat or different?
	Q16	For the rest of this section, your answers will count. The example you're going to hear is the same kind as the last two you tried. Listen to the next two sets of sounds and choose your best answer. Are the sets of musical notes the same beat or different?
Q17, Q19, & Q20	Q17 – Q25	Listen to the next two sets of sounds and choose your best answer. Are the sets of musical notes the same beat or different?
3 – Contours	Q26	Let's practice a couple of examples for the next group of questions. Play each arrow button. You will hear two sets of musical sounds. Choose the word that explains what you hear. Are the sets of musical notes the same or different? If

		you can't tell which answer is right, just choose I don't know.
	Q27	This example is the same kind as the last one. Listen to the next two sets of sounds and choose your best answer. Are the sets of musical notes the same or different?
	Q28	For the rest of this section, your answers will count. The example you're going to hear is the same kind as the last two you tried. Listen to the next two sets of sounds and choose your best answer. Are the sets of musical notes the same or different?
Q30, Q32, & Q38	Q29 – 37	Listen to the next two sets of sounds and choose your best answer. Are the sets of musical notes the same or different?
4 – Composers	Q38	Let's practice a couple of examples for the next group of questions. Play each arrow button. You will hear two short musical examples. Choose the word that explains what you hear. Do the pieces of music sound like they were written by the same or different people? If you can't tell which answer is right, just choose I don't know.
	Q39	This example is the same kind as the last one. Listen to the next two musical examples and choose your best answer. Do the pieces of music sound like they were written by the same or different people?
	Q40	For the rest of this section, your answers will count. The example you're going to hear is the same kind as the last two you tried. Listen to the next two sets of sounds and choose your best answer. Do the pieces of music sound like they were written by the same or different people?
Q41, Q43, Q44	Q41 – 49	Listen to the next two musical examples and choose your best answer. Do the pieces of music sound like they were written by the same or different people?
5 – Styles	Q50	Let's practice a couple of examples for the next group of questions. Play each arrow button. You will hear two short musical examples. Choose the word that explains what you hear. Do the pieces of music sound like they were written in a same or different kind of style? If you can't tell which answer is right, just choose I don't know.
	Q51	This example is the same kind as the last one. Listen to the next two musical examples and choose your best answer. Do the pieces of music sound like they were written a same or different kind of style?
	Q52	For the rest of this section, your answers will count. The example you're going to hear is the same kind as the last two you tried. Listen to the next two musical examples and choose your best answer. Do the pieces of music sound like they were written in same or different kind of style?

	Q53 - 61	Listen to the next two musical examples and choose your best answer. Do the pieces of music sound like they were written in a same or different kind of style?

Appendix D: Papers Written by Dawn Mitchell White Since Beginning Her Doctorate.

	Title of the Paper	Year
1	Learner-Centered Music Education: A path forward for twice-exceptional students	2023
2	Educating students with disabilities in music class: An aptitudes focus	2022
3	Twice-Exceptional music education: Best practices and a conceptual model	2021
4	The path of the music education student: A Daoist comparative philosophy	2021
5	The theories and tests of Edwin E. Gordon: Pivotal works in the development of musical measurement	2021
6	The theory of musical meaning: How do tension and release contribute to affective behaviors in the listener?	2021
7	Musical Discrimination and Skills Task: The development of an aptitude instrument for elementary music students	2021
8	On the Spectrum: the musical experiences of two young adult brothers	2020
9	The Milky Way and Golden Ratio: Juxtaposing Metaphorical Model Formulae with multi-generational musical diversity	2020
10	A review of the effects of rhythm on communication and emotional processing in Parkinson's disease	2020
11	Parkinson's disease: A review of rhythmic interventions	2020
12	Autism, emotion, and musical elements: A review with implications for music educators	2020

Note. This list accounts for Dawn Mitchell White and her works as of 6/8/2024.

About the Author

Dawn R. Mitchell White earned her doctorate at the University of South Florida with a concentration in Music Education and a cognate in Instrumental Conducting. Dr. Mitchell White maintained a private studio for thirty-one years (beginning in 1995), taught in the music classroom for 24 years (starting in 2000), and for sixteen of those years (2002 – 2018), was the sole music director at the Center for Education School of the Arts and Sciences as their band director, orchestra director, music theater director, K-4 music teacher, jazz band director, modern band facilitator, chamber music facilitator, and music theory and history educator. Most importantly, Dawn is a devoted wife (to Ron for 33 years) and a proud mom to three grown sons with autism. They inspire her to find new and exciting ways to share music with disabled learners and find "common ground" to enjoy music together. She still believes in the purity of enjoying music "just for the music's sake." However, she also views music as a "healing modality" since it offers mental, emotional, and physical benefits from listening and performing music. She has written papers about how music can be therapeutic for people with learning, developmental, and physical disabilities. She has spoken at multiple conferences about music's impact on students with autism, including the International Symposium for Music Education (ISME) in Brisbane, Australia; the ABLE Conference in Boston, at the Berklee College of Music; and the College Society of Music National Conference in Miami, Florida. She has also pioneered a conceptual model for teachers working with twice-exceptional music students.