

2022

Assessment Literacy of Middle School Music Educators

William Gordon Daines
Walden University

Follow this and additional works at: <https://scholarworks.waldenu.edu/dissertations>



Part of the [Curriculum and Instruction Commons](#), and the [Music Pedagogy Commons](#)

This Dissertation is brought to you for free and open access by the Walden Dissertations and Doctoral Studies Collection at ScholarWorks. It has been accepted for inclusion in Walden Dissertations and Doctoral Studies by an authorized administrator of ScholarWorks. For more information, please contact ScholarWorks@waldenu.edu.

Walden University

College of Education

This is to certify that the doctoral dissertation by

William Gordon Daines

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee

Dr. Mary Howe, Committee Chairperson, Education Faculty
Dr. Amy White, Committee Member, Education Faculty
Dr. Tammy Hoffman, University Reviewer, Education Faculty

Chief Academic Officer and Provost
Sue Subocz, Ph.D.

Walden University
2022

Abstract

Assessment Literacy of Middle School Music Educators

by

William Gordon Daines

MM, University of Nevada, Las Vegas, 2007

BM, George Mason University, 2004

Project Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

May 2022

Abstract

In a school district in the Southwestern region of the United States, middle school music educators did not participate in district-wide initiatives to improve teacher assessment literacy, which encompasses the assessment knowledge and practices of teachers in support of student learning. The educators varied widely in their grading and assessment knowledge and practices. The purpose of this qualitative explanatory case study was to explore the middle school music educators' perspectives of their assessment knowledge and practices in the classroom and how they used assessments to evaluate students' knowledge and skills in music. Xu and Brown's teacher assessment literacy in practice framework underpinned this study. Semistructured interviews and assessment example documents from twelve licensed middle school music educators who taught band, choir, or orchestra were analyzed using a priori, open, and axial coding strategies to derive themes. Findings showed that participants assessed student musical performances by aligning the assessment purpose to the type of assessment and grading practices, and they used grading knowledge and feedback practices to improve student performance. Based on these findings, a white paper was developed with three recommendations for music educators and district leaders. Implementation of the recommendations may contribute to positive social change through the development of music teachers' assessment literacy, which could improve their teaching practice and increase student learning.

Assessment Literacy of Middle School Music Educators

by

William Gordon Daines

MM, University of Nevada, Las Vegas, 2007

BM, George Mason University, 2004

Project Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

May 2022

Dedication

I dedicate this project study to my family.

Table of Contents

List of Figures.....	v
Section 1: The Problem.....	1
Introduction.....	1
The Local Problem.....	4
Rationale	5
Evidence of the Problem at the Local Level.....	5
Evidence of the Problem From the Professional Literature.....	8
Purpose Statement.....	8
Definition of Terms.....	9
Significance of the Study	9
Research Questions.....	11
Review of the Literature	12
Conceptual Framework.....	12
Review of the Broader Problem.....	18
Implications.....	49
Summary	50
Section 2: The Methodology.....	51
Introduction.....	51
Research Design and Approach	51
Participants.....	55
Sampling Strategy.....	55

Procedures for Gaining Access to Participants	56
Establishment of the Researcher-Participant Relationship	57
Protection of Participants' Rights	58
Data Collection	59
Data Collection Instruments	60
Sufficiency of Data	61
Processes for Collecting and Recording Data.....	63
Systems for Keeping Track of Data.....	64
Role of the Researcher	65
Data Analysis	66
Evidence of Quality	67
Treatment of Discrepant Cases	68
Data Analysis Results	68
Theme 1	69
Theme 2	75
Summary of Findings.....	83
Project Deliverable.....	89
Conclusion	90
Section 3: The Project.....	91
Introduction.....	91
Description and Goals.....	91
Rationale	92

Review of Literature	93
White Papers	93
Alignment of Grading With Assessment Purpose	95
Grading of Nonmusical Achievement	98
Use of Rubrics for Assessing Musical Performances	100
Project Description.....	103
Needed Resources and Existing Supports.....	103
Potential Barriers and Solutions.....	104
Proposal for Implementation and Timetable.....	105
Project Evaluation Plan.....	105
Project Implications	106
Implications for Local Community.....	106
Implications for Global Community	106
Conclusion	107
Section 4: Reflections and Conclusions.....	108
Project's Strengths and Limitations.....	108
Strengths	108
Limitations	109
Recommendations for Alternative Approaches	110
Scholarship, Project Development, and Leadership and Change	111
Scholarship.....	111
Project Development.....	111

Leadership and Change.....	112
Self-Analysis.....	113
Reflection on the Importance of the Work	116
Implications, Applications, and Directions for Future Research.....	116
Conclusion	119
References.....	120
Appendix A: The Project	141
Appendix B: Permission to Use the Teacher Assessment Literacy in Practice	
Illustration.....	158
Appendix C: Interview Protocol.....	159
Appendix D: Document Protocol.....	163
Appendix E: A Priori Codes and Examples.....	164
Appendix F: Open Codes and Examples	166
Appendix G: Axial Codes and Examples	171
Appendix H: Themes	172

List of Figures

Figure 1. Teacher Assessment Literacy in Practice (TALiP) Framework..... 17

Section 1: The Problem

Introduction

In 2009, the U.S. federal government announced the Race to the Top (RTTT) initiative with the goal to increase the accountability of teachers. State departments of education were granted monetary incentives to develop policies that held teachers accountable for student growth and achievement. At the time this study, state curriculum planners and assessment developers used RTTT as a foundation to create teacher evaluation programs that encompass student achievement data and teacher ratings from formal observation (Robertson-Kraft & Zhang, 2018; Wesolowski, 2015b). In some states, regulators have also used student achievement data to determine teacher pay, commonly called merit pay, causing anxiety among some teachers (Gilbert, 2016; Robertson-Kraft & Zhang, 2018). Teachers are required to hone their skills and align their teaching and assessment practices with best practices to qualify for merit.

Policy makers' and administrators' decisions about assessment are a critical factor in teachers' ability to help students grow academically; they affect the planning of future classroom instruction, the types of assessments administered, and student learning. Teachers' knowledge, beliefs, and practices related to assessment encompass their assessment literacy (AL). AL is defined as "the skills and knowledge teachers require to measure and support student learning through assessment" (DeLuca et al., 2016a, p. 248). AL is expected to fit into all aspects of professional education culture, teacher identity, and classroom practices (Coombs et al., 2018). AL is pertinent to teachers of all subjects, including music, physical education, and art, even though assessment in these subjects

has fewer implications on a school's rating or standardized assessment scores.

In this study, I sought to determine whether middle school music educators possess the AL needed for RTTT initiatives. Part of the RTTT initiative is linking student achievement data with teacher evaluations. In some states, like Florida, teachers measure the academic growth of students; these assessments, in turn, are included in the formal evaluations of the teachers (Le & Guo, 2021; Perrine, 2013). Music educators, therefore, need knowledge about assessment types, assessment data, and the application of assessment in the classroom. Teachers can promote student learning, empower the individual learner, and have a positive effect on student achievement when effective assessment practices are employed (Christoforidou & Kyriakides, 2021; Christoforidou et al., 2014). With greater AL, music educators may be better prepared to meet these RTTT requirements.

In a large mixed suburban and urban school district in the Southwestern region of the United States, many middle school music educators meet five to six times daily with more than 50 students in each class. Tracking the individual growth and achievement of each student is a challenge with multiple classes and large class sizes. Having correct knowledge of assessment and effective assessment practices may help music educators to facilitate the success of a large student population (Christoforidou & Kyriakides, 2021; Christoforidou et al., 2014).

Administrators at many school districts across the United States have assessed the progress of music programs at the group level with annual district and state festivals, but there are not adequate procedures in place to assess the individual student (Crochet &

Green, 2012; Perrine, 2016). There appears to be a gap in practice in music classrooms between the development of valid and reliable assessment methods and the assessments that middle school music educators actually create and use in classrooms (Myers, 2021; Pellegrino et al., 2015). One example of assessment use is from St. Pierre and Wuttke's (2017) descriptive quantitative study. The authors found that the assessments most often used to evaluate a music student's achievement were based on nonmusical performance and did not include an assessment of the student's achievement in musical knowledge or skill. This example is evidence of the support middle school music educators need in developing their AL.

Not only is AL pertinent to student performance ratings, but it also affects teachers' performance ratings. Each state is taking a different approach to linking student scores with teacher evaluations as part of RTTT. In Nevada, music educators use classroom assessment data to show student growth (Nevada Department of Education, 2017). A portion of the music educators' evaluations is based on the results of the classroom data. In addition, another portion of the music educators' evaluations is based on how students perform on standardized test scores in mathematics and reading. In New York, music educators are required to rate each student at the beginning and end of the year on the student's musical skills on a scale from 1 to 4 (Perrine, 2013). These scores are used to show growth at the end of the year and serve as 20% of teachers' annual evaluation (Perrine, 2013). These examples suggest that teachers, including those who provide instruction in music, need an appropriate level of AL to increase student achievement and guide classroom instruction.

The Local Problem

The problem addressed in this study was minimal AL of middle school music educators in a local school district. The site for the case study was a school district in the Southwestern region of the United States. The participants for the study were 12 middle school music educators. The middle school music educators in this district are varied in their AL creating issues in their assessment practices, according to a secondary fine arts coordinator. This gap in practice has hindered the district's music educators from assessing student learning and developing strategies to improve classroom instruction. Findings from this case study may describe music educators' understanding of AL across the United States.

Researchers in the field of music education have discovered minimal uniformity and consistency among music educators in their assessment and grading practices (Pellegrino et al., 2015; Russell & Austin, 2010; St. Pierre & Wuttke, 2017). Two major grading practices among music educators are the evaluation of musical achievement (knowledge and skills) and the grading of nonachievement tasks (student's attendance and behavior; Myers, 2021; Pellegrino et al., 2015; Russell & Austin, 2010). Teachers complete written assessments to measure students' knowledge and performance assessments to assess students' skills. However, music teachers have typically not received training on how to make or create their own assessments (Pellegrino et al., 2015). Teachers' minimal knowledge and use of music standards in assessments have resulted in invalid and unreliable assessments (St. Pierre & Wuttke, 2017). There was a need for music educators to use valid and reliable assessments, effective grading

practices (summative assessment), effective formative assessment practices, and standards to guide assessments.

Rationale

Music programs, school districts, and state music associations have assessed music performances for many years. Contest and festival ratings, which are summative in nature, include large group and solo performances and are commonplace in many school music communities around the United States (Springer & Bradley, 2017). However, in the classroom setting, music educators tend to use more subjective grading systems and, at times, do not grade based on musical achievement (Pellegrino et al., 2015).

Evidence of the Problem at the Local Level

At the beginning of the 2017-2018 school year, the leadership for the target school district initiated a balanced assessment system (BAS) for both formative and summative assessments, as noted on the website of the district's assessment, compliance, and monitoring division. District leaders designed this system to improve classroom instruction and student performance by collecting more than one set of data at multiple times throughout the school year. Music educators were given a specific time line of when assessments should occur and examples of each assessment. The formative assessments were grouped as (a) universal screeners, (b) diagnostic assessments, (c) progress monitoring and informal classroom-based assessments, and (d) district interim assessments. The examples given for each type of formative assessment are specific tests or test services and included (a) AIMSweb, (b) WIDA-ACCESS, (c) CORE Phonics Survey, (d) iReady Stands Mastery, and so forth; however, none of these examples

include any references to music assessment test services or standards. There were general examples of formative assessments in the progress monitoring section that could be used in any classroom; these included (a) self and peer assessments, (b) checklists, (c) observations, (d) rubrics, and so forth. The summative assessments of the BAS are large-scale, district-wide assessments and large end-of-unit assessments. These included (a) criterion referenced tests, (b) end-of-course exams, (c) writing proficiency exams, (d) semester exams, (e) the Scholastic Aptitude Test, and (f) the American College Testing.

The role of teachers is partially defined in the BAS document. In each assessment type, classroom teachers are listed as some of the persons responsible for the assessment, however, what they are responsible for is not clear. In the assessment cycle of planning, the classroom teacher has varying degrees of responsibility for creating and administering the assessment and recording and reporting the results, according to the website of the district's division that focuses on assessment, compliance, and monitoring. For example, in the universal screeners section there are only premade assessments from outside sources listed. For those teachers who use the premade assessments, the responsibility for assessment development is low. However, for the music educator there are no premade screeners made and music educators must create their own; therefore, the responsibility for the music educator is greater than that of the teachers who can use the premade assessments. The potential problem with having to create assessments is that it relies on the teachers' AL and possession of enough time to create a valid and reliable assessment.

This problem was also evident in the diagnostic assessments and district interim assessments. In the summative assessments, music educators would be responsible for the

creation of their semester and final exams. The need for music educators to create their own assessments was in fact addressing the larger matter of teachers' AL. The quality of the assessments depends on the AL of the teacher, and the data gained from assessments depend on the quality of the assessments. Hence, the AL of the teacher is crucial in developing a valid and reliable assessment. Those teachers who used the premade assessments are at an advantage over those teachers who must create their own, and the vulnerability of the assessments that are teacher made are high because the quality and value of the assessment hinges on the teacher's AL. The purpose of the BAS was to collect multiple types of data from every student to monitor progress and guide classroom instruction. However, if the AL level of teachers is not sufficient to collect the right data or analyze it appropriately, then the instruction that teachers provide may not be appropriate. Because the BAS alone may not produce the results the school leaders desire, it is necessary for teachers to have adequate AL to create and analyze assessment data in a correct and efficient manner.

A closer look at assessment training opportunities for district teachers further established the problem. When searching school district-sponsored professional development website for courses using the term *assessment*, I found 80 courses. Many of the courses focused on school leaders administering school wide summative assessments. There were a few subject-specific assessment courses for science and mathematics teachers to better their assessment skills. There were no courses available for music educators to improve their AL. This situation presented a disconnect between what the school leaders wanted music educators to do in the classroom and what they offered in

terms of training to music educators to meet their expectation. These documents demonstrate that there was minimal training for music educators in this school district to adequately prepare them for assessing students in alignment with the BAS.

Evidence of the Problem From the Professional Literature

Elsewhere in the field of music education, researchers have investigated problems like those in the school district study site. St. Pierre and Wuttke (2017) discovered that most of the music educators they surveyed were not familiar with standards-based grading and those who did use it in the classroom only used it because they were required to by administrators. The work of Wesolowski et al. (2015, 2016) and Springer and Bradley (2017) also indicates that the problem of inadequate AL among music educators was widespread. They investigated problems in music education AL including rubric creation, rater fairness, and validating rubrics to counteract the rater unfairness seen in research. Denis (2018) also shared this concern about what happens when music teachers have less than developed AL and how this can create learning problems for students. Because of minimal knowledge of validity or other issues related to assessment, teachers may not properly assess students, which may also have a negative impact on instruction, Denis noted. The problems of assessment practice in the target school district are like the problems being researched in the field of music education. Further investigation was needed to illuminate the issues, concerns, and AL of middle school music educators.

Purpose Statement

The purpose of this qualitative explanatory case study was to explore middle school music educators' AL in the areas of assessment knowledge and practices based on

Xu and Brown's (2016) teacher assessment literacy in practice (TALiP). The exploration of the phenomenon may bring about insight in the level of AL this population. In this part of Section 1, I described the problem and the specific gap in practice among middle school music educators. In subsequent subsections, I will present the research questions (RQs) and describe the conceptual framework I used to interpret the data and answer the RQs. The outcomes of this study may bolster music educators' assessment knowledge and practices and in turn improve classroom instruction and student achievement.

Definition of Terms

AL is the measurement of an educator's knowledge and practices of assessment (Xu & Brown, 2016). It also includes the educator's decisions within assessment and the context in which assessments occur, including outside influences (Xu & Brown, 2016). AL is complicated and difficult to measure (Baas et al., 2015), however, it can be used to identify what educators are doing in classroom assessments (Xu & Brown, 2016). An educator's AL can increase through reflection, participation in assessment activities, and work with other educators. It can change over the career of an educator ultimately defining the educator's assessor identity (Xu & Brown, 2016).

Significance of the Study

The results of this study can provide insight into the current assessment practices of middle school music educators and their beliefs and attitudes associated with their use of formative and summative assessment. The results can potentially contribute to the field of assessment and increase school leadership's awareness of the need to strengthen assessment practices. Currently, some middle and high school music teachers use non-

music-related criteria to assess student performance, which removes the focus from musical achievement and development (Myers, 2020; Pellegrino et al., 2015; Russell & Austin, 2010; St. Pierre & Wuttke, 2017). This study can provide descriptive details about the current state of assessment knowledge and practices of middle school music educators. This could include what those practices are, the resources available to teachers, and the perspectives of teachers surrounding their assessment practices.

Ultimately, the students will be the ones to potentially benefit the most from the findings of this study. The benefits of school music programs for students are not one dimensional, with benefits including many facets of life. Participating in music programs benefits students academically, cognitively, socially, and emotionally (National Association for Music Education [NAfME], 2014). Students who participate in music programs score better on tests, graduate from and attend school at higher rates, develop better relationships with others, and demonstrate an increased cognitive development (NAfME, 2014). Music programs provide students the skills to become their best selves and in turn benefit the community where they live. Students who participate in school music programs have higher mathematics proficiency and higher Scholastic Aptitude Test scores and perform better on tests measuring temporal-spatial ability (NAfME Education, 2014). Schools with music programs have a higher graduation and attendance rate than schools without music programs (NAMM Foundation, 2018). Parents and students also believe participating in music programs helps students to be more creative (NAMM Foundation, 2018). Other benefits of student participation in school music programs include positive family relationships and positive behaviors of youth (Dell,

2014; Droe, 2014; Guhn et al., 2020; Miksza, 2007). This study may promote positive social change by furthering teachers' knowledge of how they can create a more meaningful and successful music program. When students achieve at a high level, they will want to continue in school music programs and receive the benefits these programs provide (NAMM Foundation, 2018). The study was the basis for a white paper (see Appendix A) that includes recommendations that school leaders can implement to bolster music educators' AL. This study may help teachers to provide more meaningful feedback, effective lessons, and precise learning targets, which may lead to greater overall musical achievement among students.

Research Questions

Middle school music educators' knowledge and practices in AL constituted a gap in the literature that required further investigation. Research has shown how this population struggles with knowledge of, and consistency in regard to, assessment and grading practices (Atjonen, 2014; Myers, 2020; St. Pierre & Wuttke, 2017). In the school district that served as the project site, there was little or no support for middle school music educators' AL in knowledge and practices, yet these educators needed to show their student growth and achievement over the course of the year to their supervisors. Without the needed level of AL, these educators may not be prepared to complete all steps of this process proficiently. I developed the following RQs to focus this case study of the AL of middle school music educators in the district:

RQ1. What are middle school music educators' perspectives of their assessment knowledge?

RQ2. How do middle school music educators use assessments to evaluate students' knowledge and skills?

Review of the Literature

In this subsection, I review studies that relate to the study topic. The discussion will encompass the conceptual framework, background of the problem, and current discussions in the fields of assessment and assessment in music as related to the problem. Xu and Brown's (2016) TALiP provided a framework with which to explore middle school music educators' assessment knowledge and practices. The framework informed the development of the study purpose and RQs, literature review, data collection and analysis, and interpretation and reporting of the findings. The studies included in the literature review support the purpose of the study and provide an understanding to the problem of AL of middle school music educators. As I discuss, the unique nature of music classrooms requires that assessments not only be grounded in best AL practices but also that their results be applicable to music educators.

Conceptual Framework

I used Xu and Brown's (2016) TALiP conceptual framework as the framework with which to understand middle school music educators' AL. In developing TALiP, Xu and Brown drew from assessment standards and 100 peer-reviewed studies that focused on teacher assessment skills from 1985 to 2015. Their method for finding peer-reviewed studies included searching databases using keyword search terms such as *teacher*, *AL*, *assessment competence*, *professional development in assessment*, and *assessment expertise* as well as locating articles in educational assessment journals. They also

reviewed the assessment standards of the National Council of Measurement in Education, American Federation of Teachers, and the National Education Association. A requirement of the methodology in the literature they reviewed was that teachers were listed as the main targets or participants and AL was a theme. There was no mention about the exclusion of literature due to content or the level taught by the teacher. The authors used a keyword coding process and divided the articles into three main themes and 10 subthemes in the creation of the framework. The three main themes were (a) knowledge and skills within AL, (b) assessment education and its relationships with various mediating factors, and (c) contextual considerations of AL. The subthemes were (a) knowledge base; (b) measurement of teacher AL; (c) AL measurement validation; (d) assessment courses; (e) assessment training programs and resources; (f) the relationship among assessment training, teacher conceptions of assessment, and AL; (g) teacher assessment training needs and self-reported efficacy; (h) macro and micro contexts; (i) teacher's identity as assessor; and (j) understanding and development of AL in practice.

The purpose of the TALiP framework is to show school leaders how to support teachers' AL and for current teachers to see the process by which AL is increased (Xu & Brown, 2016). The three levels of mastery are (a) assessment knowledge, (b) perception of how assessment should be, and (c) one's identity as assessor (Xu & Brown, 2016, pp. 151-153). I aligned the RQs for this study with the first two levels of TALiP: mastery of music educators' perception of their assessment knowledge and assessment practices used to evaluate students' knowledge and skills.

Xu and Brown (2016) sought to connect the fields of educational assessment and

teacher education to redefine teacher AL. The foundation component is the teacher's knowledge regarding (a) content, purpose, and methods of assessment; (b) grading; (c) feedback; (d) peer- and self-assessment; (e) interpretation of assessments; and (f) assessment ethics. In the classroom, music educators would display this knowledge through an understanding of music standards and knowledge about various types of assessments. The music educators would be using standards-based grading and multiple types of assessments based on those standards. Next in the sequential order are teachers' guiding framework and conception of assessment. These two components have a causal relationship, meaning the teacher's beliefs about assessment shape the teacher's ideas, or conceptions, about the role of assessment in the classroom and for the learner. An observer looking for evidence of a music educator who understands the relationships among these components will see assessments planned that evaluate the learners' depth of knowledge. Over time, the music educator can increase the rigor of the assessments to better understand the needs of the learner. The music educator would match the assessment to the learners' needs and teach with the belief that all students can learn.

External forces create limits for teachers in classroom assessment. For example, music educators in the study district may experience mandated financial limitations, an imposed festival schedule, requirements of administrators, or large class sizes. Xu and Brown (2016) accounted for these external forces in their framework and labeled these components as *macro social-cultural* and *microinstitutional contexts*. Both components are connected as teachers often must make compromises in their assessment practices because of limitations outside of the classroom. Music educators in the target district are

not in control of the amount of funds they receive because both local administrators and state lawmakers institute decisions that limit the funds and resources available for assessment.

Some music educators have technology to assist the assessment of students' note and rhythm reading skills. This can be a time-saver for a music educator with large class sizes, but the technology costs can be too much for some schools, which limits music educators from using this time-saving technology. In addition, some music educators might have requirements from their school administrators to implement school-wide programs that are difficult to fit into music classrooms. Any school-wide writing across the curriculum requirement would be difficult in a music classroom. Often there are no desks for the students to write on, the teacher has no training in assessing writing, and time is taken away from the music learning objectives. These types of external forces can create limitations for music educators as they plan their assessment tasks.

The teacher learning component relates to the themes of (a) assessment training programs and resources; (b) the relationship among assessment training, teacher conceptions of assessment, and AL; and (c) teacher assessment training needs and self-reported efficacy (Xu & Brown, 2016). This component is an important step to increase the effectiveness of teacher AL. This framework helps teachers and school leaders understand the relationship between training, resources, teachers' conceptions of assessment, and teacher-reported efficacy in assessment. For this reason, the framework was appropriate for this exploratory case study. By using it, I addressed RQ1, which concerns music educators' perception of their assessment knowledge, and then connect

the findings with RQ2 to understand the assessment choices of middle school music educators. Xu and Brown (2016) emphasized that without teacher learning teachers will continue to use out-of-date practices or less effective assessment methods. They suggest reflective practice and participation in community activities as the two main ways teachers learn. The final component of the TALiP is assessor identity (re)construction. The (re)construction occurs throughout a teacher's career. The authors argued that there is no real ideal assessor identity; instead, there is an individual identity created by each teacher through reflection and the addition of others' perspectives.

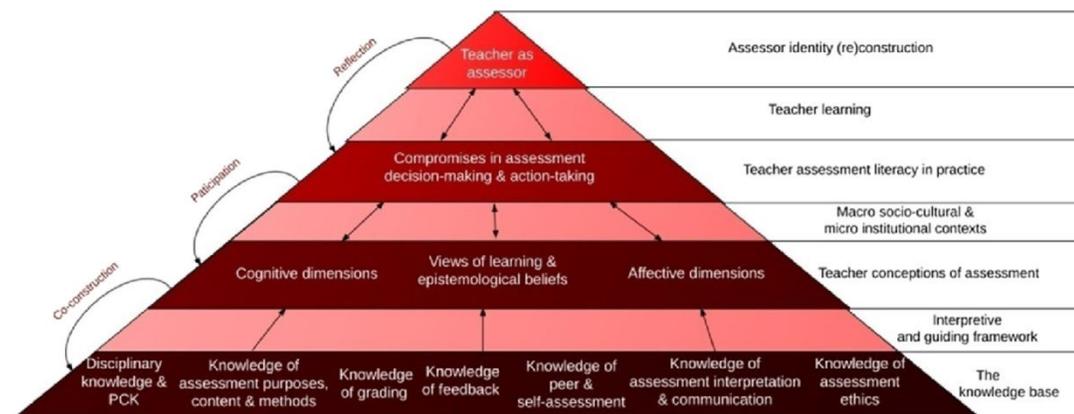
Xu and Brown (2016) observed that teachers' assessment skills and knowledge are dynamic and cyclical. Teachers may understand some aspects of AL and apply these in practice but neglect other assessment principles. However, over time, as teachers' knowledge increases about assessment, the number of omitted principles will decrease, and the teachers' AL will increase. The authors recognized that teachers' learning of assessment knowledge and skills is never static and is a continuous learning experience for teachers. Xu and Brown acknowledged (a) reflection, (b) participation, and (c) co-construction and as part of the process of gaining new knowledge in AL through experience and collaboration. Throughout the framework, the authors recognize how teacher AL practices improve and educational context change occurs. This process is directly related to real-life application, wherein teachers recognize their own strengths and weaknesses (reflection), work to improve their assessment knowledge and practice (participation), and team with other teachers to produce assessments that measure students' knowledge and skills (co-construction). This cycle is repeated throughout

teachers' educational careers. The framework's dynamic and cyclical nature evidence its accurate reflection of reality. The cyclical design can help teachers understand that continual learning and growth are necessary, regardless of the content assessed. The framework is not subject specific, and although music is unique when compared with more traditional subjects, the same assessment principles apply. The AL of middle school music teachers can therefore properly be studied using Xu and Brown's framework.

Figure 1 provides an illustration of the TALiP framework.

Figure 1

Teacher Assessment Literacy in Practice (TALiP) Framework



Note. From “Teacher Assessment Literacy in Practice: A Reconceptualization,” by Y. Xu and G. T. L. Brown, 2016, *Teaching and Teacher Education*, 58, p. 155

(<https://doi.org/10.1016/j.tate.2016.05.010>). Copyright 2016 by Elsevier Ltd. Reprinted with permission (see Appendix B).

The TALiP can be used to view middle school music educators' mastery levels of multiple components. In this study, I used the components of assessment knowledge base and AL in practice to answer the RQs. I explored participating music educators'

perspectives of their assessment knowledge and their use of assessments to evaluate students' knowledge and skills. The interview questions were also based on the TALiP framework and were the lens with which the data were analyzed and interpreted.

Review of the Broader Problem

Literature Search Strategy

To understand the current discussion about AL and assessment in music classrooms, I conducted a literature review using Walden Library resources and Google Scholar. The Education Research Complete database was used as an initial start using the search terms *assessment literacy*, *assessment AND music*, *assessment*, *NAfME*, *music standards*, and *assessment practices of music teachers*. In Google Scholar, the cited by link was used to find articles which cited the article which was useful for this review. Also in Google Scholar, the author link was helpful to find other articles which by the same researcher. This was helpful in finding researchers who spent time studying AL over a long period of time. Many of the DeLuca (DeLuca & Bellara, 2013; DeLuca & Bolden, 2014; DeLuca & Hughes, 2014; DeLuca & Johnson, 2017; DeLuca et al., 2016, 2016a, 2016b) articles and Brown (Kyaruzi et al., 2018; Panadero & Brown, 2017; Xu & Brown, 2016) articles were found using this method. Another method for finding articles connected with the AL of music educators was to use journals which contained many articles about this topic using the journal publisher's website. *Assessment in Education: Principles, Policy & Practice* and *Educational Assessment* were two journals which had assessment focused issues. Lastly, the referenced works of literature reviews and articles which had a strong connection with the problem were searched for any articles which had

promising connections. This subsection contains literature from Education Research Complete, Google Scholar, journal websites, and referenced works.

Assessment Policy: Current Context

Assessment of student learning has been a discussion point in education arena since the 1980s starting with the Nation at Risk report which led to the standards movement (Smit & Birri, 2014). From this point, the 1990s produced the Goals 2000 Act in 1994 which continued the standards focus policy (St. Pierre & Wuttke, 2017) and then an increased focus on collecting assessment data in the No Child Left Behind Act in 2002 (Richerme, 2016). The 2009 RTTT initiative then increased accountability with supporting teacher evaluation initiatives partially based on student achievement (Aguilar & Richerme, 2014). Over time, national policy makers initiated policies and incentives to increase the accountability of teachers for student test scores and thus the focus shifted from student scores to teacher accountability. Teachers' accountability for student test scores resulted in intensified pressure on teachers with evaluations that included student test scores which were a source of great stress (Gilbert, 2016).

Current literature about the use of standardized assessments to evaluate teachers show it is not beneficial for students or teachers. Heitink et al. (2016) indicated that teachers did not use best practices in their teaching when evaluations include standardized assessments. These outcomes were evident when state agencies used students' standardized test scores to evaluate teachers' annual performance. This practice resulted in limited student learning (Heitink et al., 2016). Other teachers perceived the connection of standardized test scores and teacher evaluation that inhibited their growth

of formative assessment skills (Black & Wiliam, 2018). Black and Wiliam (2018) contributed three ways to use summative assessments: (a) schools and teachers should be evaluated only on the factors which they have control over, (b) summative assessments should be developed to support learning, and (c) teachers should take responsibility for the summative process. Based on these factors, summative assessments are necessary when measuring student academic performance, while at the same time placing teacher accountability within the parameters of what teachers can control, such as, how content is taught. Currently, there are policies which burden teachers with accountability outside of their control (course content, state and district mandates, and teacher/administrative turnover) and are not in line with current research.

Connected to assessments are the standards by which the assessments are based. The development of standards has been a companion in the development of the assessment of student learning. In the early 2010s, one major discussion was about the connection of the Common Core State Standards (CCSS) created by a coalition of state governments and the Framework for 21st Century Learning (P21) developed by the Partnership for 21st Century Learning. The focus of the CCSS is college and career readiness standards and K-12 standards created for mathematics and English. Currently, 41 states, the District of Columbia four territories, and the Department of Defense Education Activity (DoDEA) have adopted the CCSS (Common Core State Standards Initiative, 2019). The P21 is a collaborative project of educators, businesses, and policy leaders. The framework P21 is based is the 21st Century Skills map which is much broader than CCSS. P21 includes English, reading or language arts, world languages,

arts, mathematics, economics, science, geography, history, government, and civics as key themes and essential to student success. These standards also place an emphasis on (a) learning and innovation skills, (b) information, media, and technology skills, and (c) life and career skills (Partnership for 21st Century Learning, 2019a, 2019b).

The CCSS does not list music as a core subject which places music in a problematic situation where the legitimacy of the subject is vulnerable and music teachers are required to defend its value to students (Gilbert, 2016). This makes music and the arts appear to be less important than other content areas which can make school leaders feel justified to neglect or remove music from school curriculum. On the other hand, P21 supports the arts as part of the necessary education of students and supports students' participation and assessment of skills in the arts. P21 also supports a wider view of assessing student achievement, not just standardized tests; that student assessments look more like performance tasks, high level cognitive demand tasks, portfolios, media presentations, and multiple steps processes (Partnership for 21st Century, 2019). These assessment types work well in music classrooms since they measure student progress over time and fit into the complexity of music assessment. This is evidence of how P21 is a better fit for middle school music educators where the CCSS lacks sophistication. Thus, using CCSS to evaluate music educators' effectiveness would lack content validity since the CCSS does not recognize music as a core subject and would base music educators' evaluations on nonmusic criteria (Gilbert, 2016). For the conflicts between CCSS and the arts just discussed, Gilbert (2016) supports the use of P21 framework is a better fit for music educators' evaluations.

As an alternative teacher evaluation, Orzolek (2018) suggested using a different framework for teacher evaluation, which includes teacher personal reflection and self-assessment as the basis for teacher evaluations and thus bypassing the discussion of CCSS and P21. In the process of suggesting this alternative framework, the author shared how “the very abstract and complex nature of learning means that the evaluation of teachers and their role in the growth of a student is never easy to appraise” (Orzolek, 2018, p. 48). To help middle school music educators and other arts teachers, the new 2014 Core Arts Standards were created to provide school leaders with standards that are measurable and help effectively evaluate arts teachers (Shaw, 2014). As a companion to the new standards, NAFME produced a teacher evaluation workbook in connection with the new standards to help music educators and school leaders (Siebert, 2014). The discussion about teacher evaluations has many facets and since they directly affect teachers, it is important to understand the debate points and the policies which influence teachers. However, as will be discussed next, music education programs are not preparing music educators to understand the policies which affects them.

Currently there are national, state, and local governments as well as professional agencies that contribute to educational policies. Middle school music educators need to understand the various assessment and teacher evaluation policies and, in addition, know how to modify their practices in the classroom to provide the desired results of student achievement (Aguilar & Richerme, 2014). But as Aguilar and Richerme (2014) found, many music education programs do not contain the policies and standards outside of music with preservice music educators. From the start of their careers, these new music

educators are behind in assessment policy knowledge and how these policies can affect their teaching and assessment practices. With the minimal knowledge about assessment policies, middle school music educators are unaware how they will be evaluated in their job performance as well as having little or no influence in shaping evaluation.

Regardless of what standards or framework music educators are used to evaluate teachers' performance, the central issue is middle school music educators' AL. Due to initiatives like RTTT, student assessment data are the data school leaders use to evaluate a teacher's effectiveness. Middle school music educators not only need to produce students who are musically independent and perform at a high level, but they also need to show student growth through their assessment efforts. However, currently middle school music educators are being evaluated not on their students' growth and achievement in music but the students' growth and achievement in other subjects.

Assessment

Assessment has been studied in many contexts with varying definitions derived from the results. Kyaruzi et al. defined assessment as “a formal or purposeful attempt to determine students' performance during and/or after a learning phase can be used formatively for improving the teaching and learning process, certifying students, placement of students in tracks, or for curriculum improvement” (2018, p. 1). Other definitions of assessment include the collecting and interpreting of data (Atjonen, 2014; Payne et al, 2019); while others include the administering and feedback as part of assessment (Christoforidou et al., 2014; Dann, 2019; Hatch, 2020; Martin, 2020). Other assessment researchers such as Black and Wiliam, define classroom assessment in their

article as “those assessments where the main decisions about what gets assessed, how the students will be assessed, and the scoring of the students’ responses, is undertaken by those who are responsible for teaching the same students” (2018, p.4). While there are differing versions of the definition of assessment, some commonalities are that teachers are at the center of assessment and that assessment is an action imposed on students.

However, Boud et al.’s (2018) case study transformed the definition of assessment through their practice theory perspective. Their aim of their case study was show that assessment needed to include the context in which the assessment took place. These researchers used 33 academics and collected data through interviews. They sought to focus on “assessment-as-practiced and how it operates” (p. 4) not only what assessments should do. Boud et al. used the framework of practice theory to guide their study focusing on the setting and time, the relationship among the players involved, the documentation created, what was produced, and the process of production. From the findings, the researchers concluded that the definition of assessment needs to also include the experience of the assessment.

Richerme (2016) agreed with Boud et al. (2018), who viewed assessment using a practice theory lens, that assessment is defined through the context or setting it is given. In her definition of assessment, Richerme (2016) added to the discussion how assessment in music is not only the measurement of students’ music-making, but the assessment process also changes the students. Students are made and remade through the assessment process (Allsup, 2015; Payne et al, 2019; Richerme, 2016), meaning that students’ musical skills are perfected through the assessment process. For example, an instrumental

music teacher may assess students' tone quality. As the assessments are administered over time the students' tone quality is refined. This is important in music environments since the curriculum is based on the performance of acquired skills which takes place over an extended period.

Another major point Richerme addressed is that many of the assessments which are administered in a music classroom are organic and not always planned nor could be planned (Richerme, 2016). Music classroom environments are rehearsal spaces where the continuing effort to refine students' musical skills and create a musical product is the focus. A music educator cannot always predict what concepts students will perform well or need further instruction. A music educator is constantly matching the conceptualized standard to the students' performance and then measuring where the student performance is lower than the standard. The music educators then place focus on those determined gaps in the student's performance. Thus, music educators are continually assessing student performance with exceptionally prompt decisions based on the recent data and then assess again. Musical assessment is the measurement and the evaluation of the musical tasks (Richerme, 2016). Musical assessments also promote student learning, give the students the opportunity to communicate musical achievement, and provide the data needed to guide instruction to improve the quality of the musical program (Myers, 2021; Wesolowski, 2015a). Assessments and measured practices can also help students become confident in their musical skills (Atjonen, 2014) and allow music making to come into existence (Richerme, 2016).

What is not mentioned in the literature is what the current AL of music educators

is or what their assessment practices are. Areas for further research could include how are music educators assessing students and what are the factors which influence the decisions of music educators. This omission in the literature justifies the need for this study to explore the assessment knowledge and practices of middle school music educators.

Formative and Summative Assessments

The two major forms of assessment are formative and summative assessments. These assessment types are defined by their purpose and not the content of the assessment. In their seminal work in 1998, Black and Wiliam defined formative assessment as activities by teachers or students which “provide information to be used as feedback to modify the teaching and learning activities in which they [students] are engaged” (p. 7). Dixson and Worrell (2016) defined the purpose of formative assessment as to improve teaching and learning and to discover where students need help. Other researchers also include the use of feedback as part of formative assessment (Atjonen, 2014; Kyaruzi et al., 2018; Martin, 2020). Almost all researchers use *assessment of learning* as a phrase synonymous with formative assessment. Formative assessment, usually informal and in low stake situations, meaning assessments which are not for a grade, have moderate to low effect on grades, and are confined to the classroom (Dixson & Worrell, 2016). Formative assessment occurs throughout classroom instruction and is developed by teachers and test publishers and covers a wide range of cognitive level questions. Examples of the different types of formative assessments can be observations, homework, self-evaluations, reflections on performance, and curriculum-based measures (Dixson & Worrell, 2016). Types of formative assessments for middle school music

educators are note reading assessments, small performance tasks during the learning process, or rhythm counting assessments

The purpose of summative assessment is the evaluation of learning outcomes and for student placement (Dixson & Worrell, 2016). *Assessment as learning* is synonymous with summative assessment. Summative assessments are more formal and occur in moderate to high level stakes, meaning the assessments have a moderate to larger influence on the grades for students and the results are discussed outside of the classroom (Dixson & Worrell, 2016). These assessments are developed by teachers or test publishers, take place after instruction, and have moderate to high cognitive level questioning. Summative assessments examples can include projects, performance assessments, portfolios, papers, in-class assessments, state tests, and national tests (Dixson & Worrell, 2016). Examples of summative assessments in music include concert performances, portfolios, or solo performances.

While formative and summative assessments are used to measure student growth and achievement, Hayward (2014) reminds us that, “There is, however, a danger that these prepositions turn into an unreflective mantra drawing attention away from the key construct – assessment is learning” (Hayward, 2014, p. 38). Assessments should be used to collect data and to inform decisions to plan the next steps for learning. Learning has taken place when the decisions have led to better curriculum alignment with learners’ thoughts and actions. However, assessment and learning occur in a complex social context which makes it difficult to evaluate. “It is within this complex framing that problems of accountability and bureaucracy sit” (Hayward, 2014, p. 38). Students are

assessed to increase students learning through data informed decisions which requires reflective practices for both teachers and policy leaders. Absent from the literature about assessment in music is the discussion about formative and summative assessments. The articles about assessment do not differentiate between the two types of assessments.

Assessment Literacy

AL is the combination of all factors which effect student assessment in the classroom. External factors like assessment policies, time limitations, and the number of available resources influence how teachers can assess student learning (Xu & Brown, 2016). Teachers' knowledge and conceptions about assessment and teachers' identity as the classroom assessor are other factors which effect assessment of student learning. AL is complicated, multifaceted, and therefore difficult to measure (Baas et al., 2015). However, Baas et al. (2015) and Heitink et al. (2016) underscored how the success of assessment is based on the assessment knowledge and skill of the educator. Gotch and French (2014) found evaluations of educators' AL to be lacking and called for further research into the development of AL measurements and how higher AL of a teacher is connected to higher achievement of students (Gotch & French, 2014). There is a need to identify what gaps educators have in AL to start finding solutions to those gaps to support student learning.

Antoniou and James (2014) concluded from their study that teachers looked on formative assessment favorably, but their assessment practices did not match their conceptions of formative assessment. The purpose of their study was to discover the current assessment practices teacher used, the decision process for using those assessment

practices, and developing a framework for the process of assessment decision making. Antoniou and James (2014) found that teachers used two types of formative assessment, (a) planned and (b) interactive assessment. Teachers did not correctly identify what assessment practices they used and often did not include students in the assessment process. Antoniou and James (2014) described the teachers' assessment practices as mechanical and not engaging the students. They also concluded the teachers favored certain assessment practices over others, such as whole group questioning and correcting, and were often at the expense of student learning, such as higher-level questioning and individual questioning. Teachers also perceived their own reporting of feedback to students as better than what happened. In a music classroom this could happen when music educators give immediate feedback to students on how to correct a musical skill. However, the music educator is placing all effect of the feedback on the memory of the student. A more effective method would be through written feedback that the student can refer to.

Thompson and Penny (2015) found similar gaps in the study of physical education (PE) teachers' AL. They interviewed 18 PE teachers from 18 different primary schools in Australia with the purpose to discover the AL of PE teachers. They found there is a gap between teacher knowledge about assessment and assessment practices as well as problems with assessment tasks validity, knowledge of the connection of assessment and student learning, how assessment relates to content knowledge, and how teacher perception relates to assessment. A major gap in AL of PE teachers is the lack of teacher knowledge about the connection between assessment and curriculum. Without connecting

assessment with curriculum, the data gathered from assessments is not a valid as to showing if students are learning or have learned. Another important outcome from both Antoniou and James (2014) and Thompson and Penny (2015) was the lack of students being involved in the assessment process. Both sets of teachers found the students absent in the assessment process including any self- or peer-assessment. The assessments were teacher decided, teacher created, and interpreted. The study conducted by Wylie and Lyon (2015) also found the inclusion of students weak in teacher AL with the addition of formative feedback being a weak area for teachers. Across disciplines educators have gaps in AL as found by the mentioned researchers.

In music, there is also evidence of music educators' lack of correct assessment practices. St. Pierre and Wuttke (2017) found a large portion of music educators who did not use standard base grading (SBG) or only used it because administrators required them. SBG is linking assessments and standards to assess students' musical achievement and providing content validity to assessments. Some music educators from the study (St. Pierre & Wuttke, 2017) used attendance, attitude, or participation as grading rather than SBG or assessing musical achievement. St. Pierre and Wuttke (2017) argued the SBG practices of music teachers should be used to identify the assessment practices of music teachers. The teachers in the study who responded they were familiar with SBG used it for a variety of reasons including, required by administrators, emphasizing learning/mastery, use of authentic assessment, and to help teacher planning and inform instruction. Seventy-five percent of those teachers who did not use SBG reported a lack of knowledge about SBG and 41.54% of those same teachers reported lack of training or

professional development about SBG as a reason for not using SBG in their classroom. These results indicate a gap in SBG grading and assessment practices by music educators.

Juchniewicz et al. (2014) elicited responses from 131 middle and high school band directors who received a Superior rating at 4 out of the last 5 years. The band directors were asked to report their most important aspects of musical rehearsals. Among the middle school music educators, there were a total of 439 items reported. Out of those 439 there were only five responses linked with assessment, (a) two middle school music educators responded the ability to properly assess was important, (b) two responded teaching students to self-assess was important, and (c) one responded with playing tests were important. The high school music educators' responses were similar with the middle school music educators. There was a total of 353 items with four responses linked with assessment, (a) three responded teaching students to self-assess was important and (b) one responded with regular and consistent assessment was important. While the items reported included all aspects of teaching, it appears there is a small disproportionate number of music educators who recognize a portion of teaching needs to include assessment to assist in student growth and achievement and pointing to music educators not giving assessment the attention needed. These studies provide evidence there is a gap in the AL of middle school music educators, especially in assessment knowledge and practice.

The Knowledge Base for the TALiP Framework

At the base of the Xu and Brown TALiP framework is assessment knowledge. Without it “there would be no standards or criteria by which the appropriateness of

assessment practice could be evaluated, potentially causing failed outcomes for teachers and students” (2016, p. 155). Under the umbrella of assessment knowledge, Xu and Brown include: (a) disciplinary knowledge and pedagogical content knowledge, (b) knowledge of assessment purpose, content, and methods, (c) knowledge of grading, (d) knowledge of feedback, (e) knowledge of peer and self-assessment, (f) knowledge of assessment interpretation and communication, and (g) knowledge of assessment ethics.

Disciplinary Knowledge and Pedagogical Content Knowledge. Heitink et al. (2016) literature review focusing on the prerequisites for assessment for learning in the classroom found 15 studies which emphasized the need of assessment being aligned with the course curriculum and that there needs to be a close link between the assessment and the curriculum; assessment cannot be an add on but integrated into the curriculum and instruction. Martin (2020), Myers (2020), and Payne et al. (2019) also insist the need for a link between curriculum and assessment. In his study of the components that effect musical assessment, Russell (2015) concludes that his model would assist music educators in diagnosing musical components when assessing students’ performances. He argued that music educators must focus on the musical components when assessing students for valid and reliable data. Then teachers can help students focus on the areas needed for improvement and help students move towards improvement in musical achievement. What is implied throughout his study is the fact that assessment needs to be linked to musical components, especially musical technique and musical expression, and how the relationship between the two effect the assessment of the performance. Without

this key foundation the assessment would be flawed from the beginning of the assessment process.

Researchers Heitink et al. (2016), Myers (2020), and Russell (2015) support Xu and Brown's (2016) framework which calls for the need for assessment to be grounded in disciplinary knowledge and pedagogical content knowledge. Through this foundation, students will know what is expected of them (Heitink et al., 2016; Russell, 2015) and foster student involved class discussions (Heitink et al., 2016).

Assessment Purpose, Content, and Methods. Xu and Brown (2016) place assessment purpose, content, and methods in the knowledge base of the TALiP for the purpose that teachers need to know why they assess, the various methods of assessment as it is related to content, and the various assessment methods. Xu and Brown (2016), DeLuca et al. (2016a), and Coombs et al. (2018) all define assessment purpose to include both formative and summative assessment. In their evaluation of AL of teachers, Coombs et al. (2018) and DeLuca et al. (2016a) found that most teachers placed a priority on formative assessment. Previously it was mentioned that formative assessment and assessment for learning are often used as synonyms, however, not all researchers agree. Jonsson, Lundahl, and Holmgren (2015) point to research which states, "formative assessment is a purpose, [assessment for learning] is a teaching and learning process" (p. 106). Others include an additional purpose to assessment, assessment *as* learning (Coombs et al., 2018; DeLuca et al., 2016a; Hayward, 2015). Assessment as learning are assessments which place the students in the role of assessor through self-reflection and

self-evaluation tasks which can include personal learning planning, document the process by which answers were given, and reflect on test preparation (DeLuca et al., 2016a).

One weakness of teachers' AL found by Gulikers et al. (2013) was that teachers often did not differentiate between formative and summative assessments and teachers did not reflect on the data collected from assessments. Gulikers et al. (2013) conducted a collaborative action research study in Dutch Agricultural Vocational Education with 11 examination secretaries from nine schools. The examination secretaries had responsibilities to oversee the assessment practices of the teachers at the school and were selected based on their expertise and knowledge of the teachers' assessment practices. Other participants of the study were teachers who were selected by the examination secretaries as exemplar of assessment practices along with three educational researchers. The question Gulikers et al. asked was, "what conceptual changes are identified as prerequisite for changing teachers' formative assessment practices to be more in line with the new outcome-based summative assessment framework?" (2013, p. 117). The researchers concluded that teachers had a narrow conception of formative assessment and viewed it as a product and not as a process part of improving instruction.

The TALiP closely connects purpose, content, and methods showing the connection between why teachers assess, what the basis of assessments are, and choosing which assessments best fit the why and the what of assessment. To better understand the teachers' understanding of the connection of assessment and student learning, the curriculum and instruction choices, and how to improve teachers' assessment capacity, Livingston and Hutchinson (2017) studied teachers' career-long professional learning to

increase assessment capacities. They concluded teachers need an individualized plan to support their understanding of connecting student learning, curriculum, instruction, and assessment. The researchers emphasized how professional learning needs to be viewed by school leaders as a process and not as an event. Additionally, they mention how teachers need mentors and continual training to increase the teachers' capacities as an assessor who connects students learning, curriculum, instruction, and assessment. Livingston and Hutchinson also warn that a one-time single initiative will most likely not provide the change policy leaders are seeking and that real change will happen through long term sustained effort of supporting teachers.

Teachers' knowledge of about assessment purpose, content, and methods as well as the connection between these assessment concepts is one of the foundational knowledge components of AL (Xu & Brown, 2016). A description of the research of assessment purpose and content was provided and a more substantial discussion of assessment methods and practices will be described at a later point in the literature review.

Knowledge of Grading. Standards based grading (SBG) is the practice of assessing students based on a set of criteria aligned with preestablished standards (Myers, 2020; Sharma, 2015) which increases the validity and reliability of student assessments (Myers, 2020; St. Pierre & Wuttke, 2017). However, as discussed previously, most music educators are not using standards-based grading and grading on nonmusical achievement criteria (St. Pierre & Wuttke, 2017). Pellegrino et al. (2015) reported that average of the responses of secondary music educators in the Southwestern region of the United States

based 60% of students' grades on nonmusical achievement categories such as attitude, attendance, and practice charts.

This gap is problematic for middle school music educators who need to not only show student achievement for teacher evaluation purposes but for also the achievement of the music student. SBG is beneficial for middle school music educators to know. SBG provides the educators with information about where students are on the continuum of learning and information for instructional planning and enables students to know their own progress (Sharma, 2015).

Feedback. The knowledge foundation of Xu and Brown's (2016) framework includes the teacher's knowledge of giving multiple forms feedback to assist student learning. One study supporting Xu and Brown is Baas et al. (2015). These researchers concluded that the feedback given to students was a predictor of the students' task orientation and planning activities. Furthermore, feedback can inform students about their strengths and weaknesses which will make them more aware of the learning tasks they need to perform. In their literature review, Heitink et al. (2016) reported nine studies which supported the need for feedback that was directly linked to learning goals. The results of Lipnevich et al. (2013) study showed how the use of rubrics, exemplars, or both rubrics and exemplars improve students' work. Lipnevich et al., along with Hatch (2020), noted how feedback through these methods without a grade can increase the level of students' cognitive thinking since the students "had to decide whether they satisfied the requirements of the assignments, figure out what exactly they did wrong, and find ways to

fix their work” (Lipnevich et al., 2013, p. 551). The evidence from these studies support how informative feedback from teachers can support students in their learning.

While feedback can be positive to help students learn, not all teachers have the same positive outlook about feedback. In a study of four elementary teachers, Antoniou and James (2014) found a gap in teachers’ knowledge in what they defined as giving feedback. Some comments were described as feedback by the teachers, but the feedback was lacking in specificity and quality. The gap was misunderstanding of what quality feedback is. In another study some teachers found it difficult to give meaningful feedback due to time constraints but found rubrics to be an effect tool to give quick formative feedback to students (Doğan & Uluman, 2017; Myers, 2010; Payne et al., 2019; Wesolowski, 2018). These studies showed how lack of assessment knowledge and minimal amount of time are two obstacles which obstruct teachers providing meaningful feedback to influence the students’ learning.

Self- and Peer-Assessment. Self- and peer-assessment are opportunities for music educators to engage students in the learning process (Xu & Brown, 2016; Zaleski, 2014) and is an important assessment practice for teachers (Kyaruzi et al., 2018). Self- and peer-assessment allow students to join with teachers in the learning process and produce better results than just teacher centered assessments (Heitink et al., 2016). Specifically, in music, peer-assessment and peer-tutoring has shown to have a positive influence on both student performance and behavior (Cangro, 2015). This evidence supports the need for middle school music educators to effectively use self- and peer-assessment in the classroom.

Valle et al. (2016) reported the use of self- and peer-assessment in a music classroom as a meaningful method of formative assessment. This type of assessment practice will help students take control of their own learning and work on closing their own achievement gaps. The researchers also reported how self- and peer-assessment is an “authentic artistic process that are apposite to music-making and important to any endeavor that involves rehearsal and redoing” (Valle et al., 2016, pp. 42-43). Through this process the students become more self-directed and self-sufficient.

While self- and peer-assessment practice can have positive effects in student learning, Antoniou and James (2014) found some teachers did not want to give up control of the assessment process to students. In their qualitative study of four primary school teachers in Cyprus, this pair of researchers found through interviews and classroom observations that all four teachers were reluctant to relinquish control of assessment to the students. Panadero and Brown (2017) found teachers have a positive view of peer-assessments but used them occasionally. In their study of 1,286 teachers in Spain, Panadero and Brown (2017) found that while teachers had a positive view of peer-assessment, teachers did not receive the training needed to support the use of this type of assessment and was viewed as requiring a lot of effort for teachers to use. Both studies highlight two obstacles teachers have using self- and peer-assessment, teacher perception and minimal training.

However, the evidence supporting self- and peer-assessment practices to increase students taking ownership for their learning and use of cognitive and metacognitive strategies (Baas et al., 2015) should encourage teachers to utilize self- and peer-

assessment in the classroom. From the students' perspective, there is also a need for teachers to teach them proper self-assessment. Hewitt (2015) found that while studying the relationship between self-efficacy and self-evaluation the "higher-performing students . . . underrated their performance ability while lower-performing students overrated their achievement" (2015, p. 307). Both the teachers and the students need guidance to fully utilize the self- and peer-assessment practice.

Assessment Interpretation and Communication. Once assessment data have been collected from the students, teachers need to interpret the data, make inferences from the data, and make choices about the next steps based on the data (Black & Wiliam, 2018). Teachers need to know how to correctly interpret the data and communicate effectively to stakeholders (Xu & Brown, 2016). This knowledge happens during methodical review of data but also "on the spot" interpretation of data as teachers interact with students in the classroom (Heitink et al., 2016, p. 56). In their literature review of AL, Heitink et al. found assessment data being used for "organizational purposes, for identifying specific weaknesses of individual students, for explaining students' thinking and problem-solving processes, and for informing teachers' pacing decisions (e.g., whether the class could move on to a new unit)" (2016, p. 57). The strength of this type of interpretation of data was summarized by Wesolowski who shared what can be done with the data collected from the assessment of big band jazz ensembles:

By highlighting the specific items that most contribute to group classification, educators can target specific performance elements that correlate to their current achievement level. This may provide a more

concrete rehearsal plan tailored to individual ensemble needs. By isolating and developing the performance criteria related to each of these items, ensembles may demonstrate the most noticeable improvement in performance achievement (2015a, p. 13)

Through correct interpretation of the data, teachers can clearly communicate the evidence of student growth and achievement, providing all stakeholders with valid and reliable data; and, therefore, eliminating any poorly or convoluted reports. Knowledge in both interpretation of data and the communication of the data are contained in the knowledge component of the Xu and Brown's (2016) TALiP framework.

Assessment Ethics. Another component of the knowledge base in the TALiP is assessment ethics. Teachers need to know and understand the legal and ethical implications of assessment (Xu & Brown, 2016, p. 156). Teachers need to know and understand the need for equity and inclusion in the classroom (Xu & Brown, 2016). Research which supports the need for teachers to understand assessment ethics includes DeLuca et al. who surveyed over 400 teachers in the United States seeking to understand Assessment Purposes, Assessment Processes, Assessment Fairness and Measurement Theory (2016). As teachers gained experience, their assessments practices became more student centered and fair. The more experienced teachers made assessments fit each student's needs which created a more equitable and fair assessment for each student (DeLuca et al, 2019; DeLuca, Valiquette, et al., 2016). In a qualitative study seeking to understand early childhood teachers' approaches to assessment, DeLuca and Hughes (2014) identified the teachers' fairness with students. They indicated all 12 early

childhood teachers interviewed shared they “explicitly valu[e] children as people and being present to children” (DeLuca & Hughes, 2014, p. 451). This type of attitude exemplifies teachers who understand the need for equity and inclusion in the classroom.

Assessment Methods and Practice

Another component of the TALiP framework (Xu & Brown, 2016) is the teachers’ AL in practice. This component, along with assessment methods, would be the visible aspect of AL. Some of the current assessment methods music educators use are portfolios (Baas et. al, 2020; Silveira, 2013; Wesolowski, 2014), rubrics (Wesolowski, 2012, 2013, 2015a, 2016, 2017; 2018; Wesolowski et al., 2017; Wesolowski et al., 2016), and standard based grading (Myers, 2021; Pellegrino et al., 2015; St. Pierre & Wuttke, 2017; Wesolowski, 2014). Another key area of focus for middle school music educators is rater fairness (Hash, 2013; Springer & Bradley, 2017; Wesolowski, 2016; 2018; Wesolowski & Wind, 2017; Wesolowski et al., 2015; Wesolowski et al., 2016; Wind et al., 2016), or how to address the bias of the assessor. In the following topics, studies which highlight each of these areas of assessment methods and AL in practice will be discussed along with the connection to middle school music educators’ AL.

Portfolios. Portfolios are used as pedagogy tool, learning tool, and assessment tool (Baas et. al, 2020; Lam, 2016) and are used across all content areas, including music. In his literature review of the use portfolios among English language learners, Lam (2016) reported how portfolios could lead students to better self-reflection and independence in writing. Blom et al. (2014) supported Lam (2016) in their research of four Australian universities in music and writing and found that portfolios allowed for

student self-reflection, peer evaluation, and essay collaboration. Lam (2016) further described in detail how portfolios can be used in assessment as learning, assessment for learning, and assessment of learning; thus, portfolios can have multiple uses depending on the phase of teaching with feedback of portfolios being a key component. Portfolios can be used as a formative assessment with the benefits in six areas: (a) process assessment, (b) authenticity, (c) integration and learner centeredness, (d) learner autonomy, reflection, and responsibility, (e) motivating, and (f) enhanced writing performance (Burner, 2014).

In the music classroom portfolios can be used to expand learning beyond performing (Baas et. al, 2020; Silveira, 2013) and documenting student learning (Baas et. al, 2020; Burner, 2014, Silveira, 2013; Wesolowski, 2014). Silveira (2013) encouraged the use of portfolios to facilitate the updated 2014 music standards: performing, creating, and responding. Portfolios in music classrooms allow for teachers, students, parents, and administrators to see progress over time of students' performances (Silveira, 2013). Students' creations can also be documented and collected using the portfolio practice. Students can collect their improvisations, composing, or arranging which will allow for self-reflection of their own creation (Silveira, 2013). Educators can use portfolios to respond to music through analysis, evaluation, relating music to the other arts, and relating music to history which will allow for students to see connections with music content and other disciplines (Silveira, 2013).

An important strength of the portfolio practice is the ability to document student learning. Educators in many states are having to document student learning and

achievement for positive yearly evaluations and the portfolio is a method which can show individual student growth (Myers, 2021; Silveira, 2013; Wesolowski, 2014). Wesolowski (2014) documented how music educators can use portfolios in the diagnostic assessment, formative assessment, and summative assessment of students' learning. The strength of the portfolio is the opportunity it gives for teachers to individualize or differentiate learning for students. They also provide music educators to collect various forms of data to improve planning and instruction (Payne, 2019; Wesolowski, 2014). While there are strengths to the use of portfolios in classrooms, music educators are the least likely to use this form of assessment (Wong, 2013, Yan, 2021). The implementation of portfolios is perceived as challenging and not suitable for a music classroom (Wong, 2013). This gap in music educator knowledge of how the practice of portfolios can increase student achievement, document student learning, provide quality feedback, and differentiate learning limits the use of portfolios in the music classroom.

Rubrics. A major problem with assessing musical performances is how to assess both the technical and expressive aspects of music (DeLuca & Bolden, 2014; Wesolowski, 2012). Rubrics offer a way for music educators to assess both aspects of a musical performance by evaluating the complexities of a musical performance in a formative way with feedback (DeLuca & Bolden, 2014). Also, many music educators have large class sizes which require efficiency and accuracy amid a heavy workload. Rubrics also offer music educators a method to score performance assessments quickly with quality feedback (Doğan & Uluman, 2017).

The creation of rubrics is critical for the effectiveness of the rubric and music

educators need to know how to select the criteria. Choosing to assess certain criteria of music will exclude or minimize other criteria important to music performance (Richerme, 2016). This can make rubric construction potentially problematic and, therefore, the rubric criteria need to be founded in standards with clear descriptions. The rubrics can then be used for feedback to students (DeLuca & Bolden, 2014; Myers, 2021; Payne et al., 2019; Smit & Birri, 2014). The use of learning goals and purpose of the assessment will also ground the rubric with the clarity and specificity needed (Wesolowski, 2012). Rubrics can have the same usefulness in all classrooms but as Williams (2015) pointed out, there is a gap among teachers using rubrics, who report they are not adequately skilled in creating rubrics. Williams (2015) argued for the use of rubrics since they can be used for both formative and summative assessments and can be anchored to curriculum standards.

Wesolowski is a researcher who has spent much of his research efforts in studying rubrics and rater fairness in musical performance assessments in both the wind band and jazz band idioms. Through his work in solo wind band (Wesolowski, 2021; Wesolowski et al., 2018; Wesolowski et al., 2017; Wesolowski & Wind, 2017), jazz rhythm section (Wesolowski, 2017), jazz big bands (Wesolowski, 2015a), and rater fairness in wind band (Edwards et al., 2019; Wesolowski, 2016; Wesolowski et al., 2015; Wesolowski et al., 2016; Wind et al., 2016), Wesolowski developed rubrics to provide music educators with an assessment tool which can be used to guide both instruction and assessments (Wesolowski, 2017). Wesolowski observed that “the development of valid and reliable assessment tools cannot only improve the teaching and learning processes but can also

provide insight into the more elusive aspects of musical behavior” (Wesolowski, 2013, p. 241). He further gave support for the use of rubrics offering nine benefits of using rubrics in a music classroom:

(a) clear levels of accomplishment by defining tangible measures of individual achievement, (b) clear indications of what students need to accomplish in the future to improve their individual performance, (c) a learner-centered approach to performing, learning, and assessing, (d) a bridge between student learning and teacher expectation, (e) versatility in adapting to meet the needs of a specific curriculum, student age, ability level, style of music, and type of ensemble, (f) a valid and reliable form of individualized assessment and documentation of teacher accountability, (g) a quantitative means for evaluating and scoring qualitative, performance based tasks, (h) a means for clearly implementing content standards and course objectives into the assessment process, and (i) valuable information for parents on their child’s progress and needs for improvement (Wesolowski, 2012, p. 38).

Rubrics are an assessment practice which can provide middle school music educators the means to provide accurate assessments in a complex and subjective discipline.

Rater Fairness. There can be difficulties with assessing musical performances due to the subjective and complex nature of music (DeLuca & Bolden, 2014; Iusca, 2014; Springer & Bradley, 2017; Vidwans et al., 2017; Wesolowski, 2012; 2018). Another problem of assessing musical performances is how to control for rater fairness or rater bias. These two problems can be solved by the same assessment practice of using quality

developed rubrics. However, the problem with rater fairness is the gap between the content in the rubric and the observed score (Wesolowski, 2016). If this is not checked, the assessment tool is assessing the rater and not the performance (Wesolowski, 2016). Incongruence in rater fairness can also lead to “unfair failing and unfair passing” (Gerritsen-van Leeuwenkamp, Joosten-ten Brinke, & Kester, 2017, p. 95) causing some students to receive a certificate but not performing the necessary skills while others could perform the necessary skills but not have the certificate. In their investigation into instrumental solo scoring, Wesolowski and Wind (2017) claimed the following factors were why there would be negative rater fairness scoring:

(a) lack of validity evidence of the functioning of the measurement instrument itself, (b) use of scoring procedures based upon Classical Test Theory that confuse raw scores for linear measures, (c) lack of empirical evidence of rater behavior, (d) lack of standards-based scoring procedures, (e) lack of rater training where clarity in the use of the instrument and knowledge of performance standards are clear, and (e) lack of rater recalibration throughout continued assessment contexts (p. 15)

These problems are difficult for middle school music educators to overcome on their own. However, there are steps that can be taken to ensure a more valid and reliable assessment such as (a) an assessment tool based on music standards or benchmarks, (b) receive training in assessment, and (c) recalibrate oneself regularly with exemplars (Wesolowski et al., 2015). The improvement of middle school music educators’ AL skills as rater fairness will provide these educators with reliable data and fair assessments of

students (Payne et al., 2019; Wesolowski et al., 2016).

Music Standards

In 1994 the national arts standards were part of the Goals 2000: Educate America Act of 1994 (Aguilar & Richerme, 2014). By the year 2000, these standards helped 49 states to develop and adopt their own arts standards (Rawlings, 2013). However, the introduction of the standards did not have the desired effect all over the country. Teachers did not feel adequate to teach or implement all the standards in the classroom. The implementation of the standards was a challenge to the many teachers due to varying circumstances such as school location, class size, resources, and community support (Aguilar & Richerme, 2014). Some studies also indicated course syllabi were changed by the standards but the teaching strategies by professors did not change (Aguilar & Richerme, 2014), showing how the standards lacked any meaningful effect.

In 2007 the Music Educators National Conference (now known as National Association for Music Education [NAfME]) surveyed music educators to see they wanted updated standards, but the results were indeterminate; however, by 2011 every person surveyed in 40 states wanted new arts standards (Rawlings, 2013). A coalition of government organizations and educational organizations, including the State Education Agency Directors of Arts Education (SEADAE). The current president of SEADAE was one of the writers of the 2014 arts standards and mentioned there were three main reasons for changing the previous 1994 arts standards: (a) the current techniques for creating in the arts has changed, (b) technology has changed, and (c) a response to the common core standards in mathematics and English (Rawlings, 2013).

Starting in 2012, the National Coalition for Core Arts Standards (NCCAS) had 60 writers divided into teams of 10-12 writing standards in the disciplines of dance, media arts, music, theater, and visual arts (Rawlings, 2013). In 2014 the new standards were released. The framework for the standards is based on the four artistic processes of creating, performing, responding, and connecting (National Coalition for Core Arts Standards, 2014). Each process has anchor standards with the performance standards being discipline specific. Within each performance standard, there are achievement standards based on grade level creating measurable and specific learning goals (National Coalition for Core Arts Standards, 2014).

For the purposes of this study, the music traditional and emerging ensembles strand will be used to understand middle school music educators' AL. Since the performance standard is measurable, middle school music educators can use these standards in both formative and summative assessments. While the conceptual framework for this study is the TALiP, the music standards middle school music educators use will also explain the purpose of assessments and the reasons behind why certain types of assessments are selected.

Model Cornerstone Assessments. The new 2014 core arts standards also include Model Cornerstone Assessments (MCA) which provide music educators with examples assessments which can be used with each artistic process and anchor standard. These MCAs are for educators to use and adapt to their circumstances and are performance assessments which will assess the way music educators teach. Tuttle suggested that music educators who are forced to make pencil and paper test are losing validity of their

assessment (Rawlings, 2013). Tuttle also asserted these assessments which “fails consequential validity, because it will change the practice of adults in the classroom, away from having students creating in the arts to preparing for the pen-and-pencil vocabulary test . . . The answer lies with rich performance/product/portfolio assessment” (Rawlings, 2013, p. 161). The MCAs provide music educators with assessments which answers Tuttle’s concern about traditional assessment methods.

McCaffrey is the current state arts consultant for the New Hampshire Department of Education, is the current president of the SAEDAE, and part of the team of writers for the 2014 Core Arts Standards (CAS). McCaffrey supported the new standards and informed the MCAs are examples of how to assess students in alignment with the standards. Together, the CAS and the MCAs will help students to see their progress and the learning goals (Payne et al, 2019; Shaw, 2014).

The advantage of the MCAs for music educators is the adaptability for educators to fit the assessments to their circumstance (Richerme, 2016). The MCAs can also be used as a standardized assessment to enable music educators to collect data on their program (Shaw, 2014) or as formative assessments to monitor growth of students (Richerme, 2016). These updated model assessments provide teachers with content valid assessments, which are adaptable, and will give meaningful data connected with student growth and achievement.

Implications

The possible outcomes from this research study may show middle school music educators’ AL deficient in some areas. With the potential to find a gap in assessment

knowledge and practices, the project was a white paper, which supports the development of middle school music educators' AL. I used the findings of the study to inform the project goals and objectives of the professional development training. The findings guided the decisions to determine which topics to discuss in the white paper.

Topics included in the white paper were derived from the findings of the study. These topics included Alignment of Grading Practices with the Assessment Purpose, Grading of Nonmusical Achievement, and Use of Rubrics in Grading Assessments. The findings from the research did inform the scope and depth of each area of emphasis.

Summary

The focus of this study was the minimal AL skills, mainly the knowledge of assessment and assessments practices, among middle school music educators in a local school district in the Southwestern region of the United States. Section 1 contained the problem with evidence at both the local level in the professional literature and the purpose of the study. It also included an explanation of the significance and two guiding RQs of the study. The conceptual framework was the TALiP framework (Xu & Brown, 2016) which describes AL skills of teachers. A thorough literature review was conducted and indicated that middle school music educators' grading and assessment practices was not at the level needed to promote student learning and achievement. Research in AL indicated that teachers' knowledge about assessment and assessments practices were low. Based on the literature, the possible implications of the study were described with the project description. In the next section, the research design, data collection procedures, and data analysis procedures will be described.

Section 2: The Methodology

Introduction

In this section, I describe the methodology that I used to investigate middle school music educators' AL with the focus on assessment knowledge and practices. This section includes a discussion of why the qualitative case study research design (Yin, 2009) was selected and other qualitative approaches were rejected. Information about the participants, data collection, and data analysis is included. I also present the data analysis results.

Research Design and Approach

The intent of this study was to investigate the assessment knowledge and practices of middle school music educators to provide insight that educational leaders can use to improve these educators' AL. As such, a qualitative approach was appropriate. Creswell (2012) listed the following reasons for using qualitative research: (a) to learn about the views of individuals, (b) to generate theories based on participant perspectives, (c) and to obtain detailed information about a few people (p. 64). Qualitative research, Creswell added, generally allows for more in-depth inquiry into, and understanding of, the problem than other research methods (Sheldon et al., 2010). The strength of the qualitative method is that it allows researchers to understand "how people interpret their experiences, how they construct their worlds, and what meaning they attribute to their experiences" (Merriam, 2009, p. 5). A major factor in middle school music educators' AL is the decision process each educator uses and why they make the decisions they make about assessment. Therefore, the decision to use the qualitative method was based on the need

for an in-depth exploration of middle school music educators' assessment knowledge, practices, beliefs, and attitudes.

I rejected quantitative and mixed methods because neither would adequately provide answers to the RQs. Quantitative researchers use a deductive approach and use the collected data to support or refute a hypothesis (Shapiro, 1973). This approach is a process designed to be objective wherein the results are placed into predetermined categories (Shapiro, 1973). Additionally, quantitative survey research can include closed-ended questions whereas qualitative surveys feature open ended questions. There was a need to explore middle school music educators' experiences around assessment. I determined that open-ended questions would provide for deeper inquiry and greater understanding of the study phenomenon (see Patton, 2002). Researchers use the mixed-methods approach when answering the RQs requires them to collect both quantitative and qualitative data or if the problem requires more than one method to fully investigate the problem (Shapiro, 1973). The RQs, the problem, and the purpose of this study did not require the use of mixed methods. For these reasons, both quantitative and mixed methods were rejected.

I used a case study design for this qualitative research. I heeded Yin's (2009) description of the scope and design of the case study. Case study methodology is an empirical inquiry in which a researcher investigates the real-life context of a contemporary phenomenon (Platt, 1992). In conducting a case study, the researcher seeks to develop a system with boundaries, which is referred to as a bounded system (Koosimile, 2002). A bounded system is described as what is "fenced in" (Merriam,

2009, p. 40); this can be an event, activity, process, group of people, or individual (Creswell, 2012, p. 465). Yin defined this bounding as the “case” and divided the type of cases into four major types: holistic (single case), holistic (multiple case), embedded (single case), and embedded (multiple case; p. 42). The case type selected for this study was a single case study. The case was the AL of middle school music educators in one urban school district, and the units of analysis were the 12 individual music educators.

The advantage of case study inquiry is that it can include more variables than data points, rely on multiple sources of evidence, and use theoretical propositions to guide data collection and analysis (Platt, 1992). This methodology also illuminates decisions or a set of decisions in organizations, programs, processes, and neighborhoods by describing how the decisions were implemented and the results of the decisions (Basu et al., 1999). Yin (2009) noted that there are four different applications for case study methodology. These are to (a) explain presumed causal links that are too complex for survey or experimental strategies, (b) describe a treatment plan and the application thereof in the conditions it took place, (c) illustrate certain topics within an evaluation in a descriptive mode, and (d) enlighten those situations in which the intervention being evaluated has no clear set of outcomes. The methodology of case study enabled me to explain in depth the complex relationships involved in the AL of middle school music educators.

I considered other research designs for the study, but I rejected them based on the problem, purpose, RQs, and conceptual framework. The type of data needed for this study were the “richly descriptive” (Merriam, 2009, p. 16) words and experiences of the participating middle school music educators, which fit qualitative methodology.

Qualitative researchers relying on the grounded theory design and use comparative analysis to develop a theory with the emphasis on inductive strategies; they refrain from testing theories (Bunch, 1998). This design provides researchers a framework with which to analyze the data, helps researchers contemplate alternative meanings, and brings clarity to the development of theories (Chen et al., 2008). I rejected this design because the RQs are not focused on a process or the development of a theory but rather on the perceptions of the middle school music educators to discover their current knowledge and practice of assessment. I considered, but ultimately rejected, the ethnographic design. Although I focused on a certain group of educators, I did not focus on the culture in which the educators work. An ethnographic approach would focus on the culture of the educators and what decisions were made based on the culture which is not the purpose of this study (Wainwright et al., 2006).

In the phenomenological design, the researcher creates meaning based on how the participants interpret their experiences (Trotman, 2006). This type of focus would not answer the RQs because the study concerned gaps of assessment knowledge or practice and the participants cannot provide information they themselves do not know. Historical design is a qualitative design in which the researcher uses past experiences to understand the present and predict future events. The RQs focus on the experiences and assessment knowledge and practices of middle school music educators. Use of the historical design would not have allowed me to answer these questions and was rejected. For these reasons, the case study methodology was selected to provide the most insight into the phenomenon.

Participants

The population for this study was middle school music educators. I drew the sample for this study from music educators who teach band, choir, or orchestra from the 59 middle schools in the local school district. Twelve participants were selected from this population to explore the phenomenon. At the time of this study, there were a total of 149 middle school music educators with 60 band music educators, 50 orchestra music educators, and 44 choir music educators in the district. The participants needed to teach at one of the local school district middle schools; hold a valid teaching license; and teach band, choir, or orchestra.

Sampling Strategy

Nonprobability, or purposeful, sampling is the preferred method of selecting participants in qualitative research (Merriam, 2009). Purposeful sampling, or what some prefer to label as *criterion-based selection* (Merriam, 2009, p.77), involves the researcher selecting participants based on a set of criteria that will most likely provide quality insight into the phenomenon being studied. The criteria for participants for this study were middle school music educators who teach band, choir, or orchestra with a valid license in secondary music. A subcategory of purposeful sampling is maximal variation sampling; this technique allows for multiple perspectives in a complex situation (Creswell, 2012, p. 307). Following maximal variation sampling, I selected 12 participants who represented three music categories: band, choir, and orchestra. The participants included four band music educators, four choir educators, and four orchestra educators from the total district middle school educators.

The purpose of this qualitative explanatory case study was to explore middle school music educators' AL in the areas of assessment knowledge and practices using Xu and Brown's (2016) TALiP framework. There is a balance needed between breadth, normally meaning a larger the sample size, and depth, normally meaning a smaller sample size (Patton, 2002). Twelve participants in this study provided a balance between diverse and similar understandings of the phenomenon AL, fit within the limitations of resources, and provided the "richness" (p. 245) of information needed in qualitative studies (Patton, 2002). Using four educators in each music discipline allowed for breadth among the music disciplines and depth within each music discipline, thus yielding more data for analysis and comparison.

Procedures for Gaining Access to Participants

To gain access to the participants, I requested permission from Walden University's Institutional Review Board (IRB) and the division that focuses on assessment, compliance, and monitoring in the local school district. The school district required an online application with the submission of a letter of intent and research application. After the submission of letter of intent to the division, a research application needs to be completed. The review board notified me within 1 week after reviewing the research application with permission to conduct my research. When the application was approved, I asked permission of the middle school principals using the letter created by the school district. From those principals who gave permission, I recruited participants for the study.

Recruitment procedures included a recruitment letter sent to all the middle school

band, choir, and orchestra educators in the school district. The recruitment letter described the purpose of the research study, the expectations of a participant, when the study took place, verification that all information about the participants and their data will be kept confidential, and how to contact the researcher. I emailed potential participants a copy of the recruitment letter, and they had 7 to 10 days to respond to the recruitment letter with a reminder email sent on Day 6.

Establishment of the Researcher-Participant Relationship

The relationship between the research and participant is complex. There is a power dynamic which lends the researcher to make precise choices to develop a rapport with the participants to encourage trust (Merriam, 2009). The researcher works on developing trust when the procedures are “systematic and rigorous” which can increase the depth of insight of information gained from the participant (Bogdan & Bilken, 2007).

The participant is a valuable respondent when he or she has knowledge and experience about the phenomenon. The relationship between the participant and the researcher during the interview process is comprised of three components: (a) the interviewing skills and demeanor of the interviewer, (b) the mindset and openness of the interviewee, and (c) how each person views the interview (Merriam, 2009). Knowing this, as the researcher, I created procedures and established an environment to help develop positive rapport and trust with the participant. The procedures helped participants to respond to given prompts and have their voices heard (Merriam, 2009). In the recruitment email and participant consent form, the participants were notified of the expectations of a participant and my duties and obligations to the participant. In these

documents the participants were informed of their protections and guarantees from myself and who to contact if they had concerns or felt these protections were violated.

As the researcher, I treated the participants with respect and discretion. I developed trust so the participants know I will keep all information confidential (Bogdan & Bilken, 2007). I spoke to the participants in a friendly and professional manner, managed my emotions, and avoided last minute changes to schedule or protocols. I was positive and clearly described the purpose of the study without any deception. I collected data without judgment, being sensitive, and respectful of the participant (Merriam, 2009).

Protection of Participants' Rights

Participant confidentiality was protected throughout the study. To protect their confidentiality, each participant was given an alpha-numeric identifier which was used to store data for easy retrieval and to report the findings. For example, a participant who is an orchestra teacher would be labeled as OR01. Prior to participation, each participant was emailed an informed consent form. The consent form included the following components: (a) the purpose of the study, (b) statement informing the participation in the study is voluntary and the participant can leave at any time, (c) the minimal risks involved with the study, (d) the potential benefits of the study, (e) promise of confidentiality, (f) who to contact with questions, (g) who to contact for questions about participants rights, and (h) consent to the interview and sharing of assessment documents. Through the consent form, the participants were informed they can decline to answer any question and can end the interview at any time without recourse or punitive actions against them. Participants were emailed the consent form to me 1 week prior to the

interview. Two participants did not send me a signed consent form and I brought a blank consent form available at the time of the interview. As a thank you, each participant received a \$10 gift card for their time in participating in the study.

All audio files, transcripts, and documents are stored on my password protected laptop and backed up using Google Drive password protected cloud storage. All data for this study will be stored for 5 years upon completion of the study and then will be deleted from the laptop and cloud storage, including the shredding of all paper documents.

Data Collection

The qualitative data collection instruments are semistructured interviews and document review. Also included in this subsection is an explanation of the sufficiency of data, how and when data will be collected, data storage procedures followed by the description about the researcher's role and potential areas of bias. Interviewing is the most common form of qualitative data (Merriam, 2009). Interviewing is needed when the data needed are not observable or when the researcher is interested in past events (Merriam, 2009). Components of the TALiP are concerned with the teachers' assessment knowledge, the compromises teachers make in assessment decisions, and practices. Attempting to observe these components did not provide the level of understanding or data needed to answer the RQs. Interviewing the participants provided the data needed relative to AL.

A second data set collected was two examples of formative and summative assessments from the middle school music educators. These formative and summative assessment documents provided support of what the participants say in the interview and

give a “snapshot” of what the middle school music educators do in the classroom (Merriam, 2009, p. 140). These assessment documents also provided insight into the assessment skills, knowledge, and practices of music educators. These standard data collection instruments for qualitative studies provided a rich wealth of information into the AL of music educators.

Data Collection Instruments

Interviews

Interviews allow for in-depth insight from participants about the RQs (Roulston, 2014) and documents are valuable source of data since they are stable and unaltered by the presence of the researcher (Merriam, 2009). The purpose of this embedded single case study was to explore middle school music educators’ AL with a special focus on their assessment knowledge and practices.

Semistructured interviews were used with 13 questions which were asked of each participant. The semistructured interview added flexibility in asking additional probing and follow up questions to gain more insight into each participant’s AL (Merriam, 2009). The interview questions were open-ended and provided for a wealth of data about the participants experiences, opinions, perceptions, and knowledge (Patton, 2002) about assessment. I created the interview questions using the TALiP framework and related literature as a guide (see Appendix C). Probing questions were also used during the interviews. These questions were used until the RQs were answered to the extent the participant could or there was redundancy in the responses from the participant.

Document Review

The second data type collected was two assessment samples from the participants. Pellegrino described two types of “music-making data,” the *process* of music-making data and the *product* of music-making data (Pellegrino, 2014). For this study, the process of music-making data was considered formative assessment, or assessment for learning, while the product of music-making data was considered summative assessment, or assessment of learning. A worksheet which contains questions pertaining to music note identification is a document which shows the process of music-making and was an example of formative assessment. An example document of the product of music-making data was the rubric used to assess a student on the performance in a concert or solo. Both assessment formats provide documentation on student learning but are at different stages of the learning process. Pellegrino suggested using these types of data points which provided insight into “topics such as pedagogy, curriculum, student learning, . . . and music teacher identity” (Pellegrino, 2014, p. 322), all of which are part of the RQs of this study. At the end of the interview, the participant was asked to provide a copy of two assessments. The documents retrieved from the participants provided an understanding of the participants’ knowledge and practices in assessments, which link the TALiP framework and the RQs. The document protocol can be found in Appendix D.

Sufficiency of Data

Qualitative data collected from middle school music educators gave insight into their perspectives about assessment knowledge and practices by telling their “lived experiences” (Miles & Huberman, 1994, p. 10) that cannot be seen by an observer

(Patton, 2002). Their words are data which are full of meaning and were a “special kind of information” which yielded a greater understanding into the gap in practice, the local problem, and the RQs (Merriam, 2009, p. 88). The participants were interviewed which provided data to answer the RQs (Miles & Huberman, 1994). The data collected were in the form of words which are intricate and multilayered. Words are “fatter” than numbers, have multiple meanings, and “require some processing” (Miles & Huberman, 1994, p. 9 & 56). The middle school music educators’ interviews yielded a multilayered data set that produced the needed amount of data to answer the RQs.

Documents “constitute a particularly a rich source of information” (Patton, 2002, p. 293). The information from this data set can supply information which cannot be observed and is an historical record of what has already taken place in the middle school music educators’ classroom. Documents confirmed the teachers’ interview data and provided insight into what assessment practices were the most important to the middle school music educators. Together, the data sets of interviews and documents provided sufficient data to answer the RQs.

Sample size is a factor in determining sufficiency of data. Patton (2002) instructed the need to choose between a smaller or larger sample size based on the needs of the study. The smaller the sample size the more in-depth data can be collected while larger sample sizes help to generate multiple meanings into the problem. The problem and RQs for this study required analysis of in-depth data. The gaps in current knowledge about middle school music educators’ assessment knowledge and practices require a smaller sample size to understand the phenomenon more fully. The data from the 12 participants

provided data which were valuable to gaining an understanding into the middle school music educators' AL in the areas of knowledge and practice.

Processes for Collecting and Recording Data

After all the necessary permissions were granted from Walden University IRB and from the school district leadership data collection occurred. The interviews were scheduled within a 4-week period. When scheduling the interviews, efforts were made to meet with the participants at their earliest availability at a safe and secure location. Participants were emailed three options for an interview date, time, and location with the option to propose other options if the ones suggested did not work. Interviews took place at the participant's school in a private room during nonteaching time. Signs were placed on the interview room door to secure the room and to avoid disruptions or unnecessary pauses of the interview. A reminder email was sent to each participant the day before the interview with date, time, and location information.

The interviews were conducted one on one during a one time 45-60 minute interview and were audio recorded on a smartphone. Prior to the first interview, a session with a colleague was arranged to rehearse the interview protocol and recording procedures. At the time of the interview, I used a script to read to the participant at the beginning and end of the interview. In the script I introduced myself and reminded the participant of the details of the study, the rights of a participant, and the expectations of the interview. The participant was also reminded of confidentiality and what steps were taken to protect all data. At the end of the interview the participant was thanked for their time, given instructions on the procedures for member checking, and made sure they have

my contact information.

During the interview, there were a set of 13 questions that were asked of each participant with additional probing questions as needed to gain further data or clarification about the participant's response. As the interviewer, listening was an important skill since this led to better probing questions and, therefore, gathering better data (Merriam, 2009). In the first questions, I sought to identify the background of the participant and then move to questions about the participant's assessment knowledge and practices. At the end of the interview the participant was asked to provide a copy of two assessments which fit the process and product of music-making.

Once the data were generated and gathered, the data were stored using the QDAS NVivo 12 Pro. In this QDAS, the audio files, the transcripts, documents, and research journal were stored. The audio files were transcribed by me using nuance dragon speech software and each document was scanned as a pdf file. Both file types were imported into NVivo and were password protected.

Systems for Keeping Track of Data

All qualitative data were collected and stored in the NVivo 12 software program which was used for data management. The project was password protected in NVivo. Yin stressed the importance of case studies have "presentable" evidentiary database for future researchers to see (Yin, 2009, p.119). To follow Yin's direction, a cloud-based storage folder is used to house all data collected to make the data easily accessible. All audio recordings, transcripts, and documents were saved in a folder and labeled using an abbreviation of the subject and participant number. For example, an interview conducted

with an orchestra teacher is labeled OR01. The data were made presentable by labeling transcripts and documents by participant pseudonym and then placed those documents in a folder tree to make the data easy to find. This cloud-based storage was password protected and would only be shared based on Walden University permission. Also included is the researcher's journal and notes which included decisions made during the development of the study and outcomes based on those decisions (Stevens & Cooper, 2009). Yin suggested the notes and journal of the researcher is some of the most valuable data of the database (2009). The collection of data in an organized form will allow outsiders to view the data, see the chain of evidence used by the researcher, and is an audit trail strategy for increasing the validity and reliability of the study (Yin, 2009).

Role of the Researcher

I have been a middle school music educator for 14 years in the school district where this study took place. I have taught at the same school my entire professional career. As a music educator I have no role of authority over any other middle school music educator and my professional role should not affect the collection of data. I know some of the teachers in the district; however, with more than 50 middle schools, I do not know all the middle school music educators. Many of the relationships I have with other middle school music educators in the district is strictly professional and in-person interaction with them may occur only a few times a year.

As an experienced teacher, I have increased in my teaching knowledge and skills in many aspects of teaching. In collaborative groups in my school and in groups of music teachers, we have discussed assessment in some detail. I have noticed a pattern of

assessment beliefs among some teachers but have not had the experience to observe any assessment practices. To control for this potential bias, I wrote unbiased RQs. The questions were “big, expansive questions” (Jacob & Furgerson, 2012, p. 4) which allowed participants to share their experiences without any influence from myself. In addition, the question topics were centered on the TALiP framework and current literature so that my opinions and beliefs are not included in the question topics and do not lead the participant to answer the questions in a way influenced by my personal bias.

Member checking was used to provide credibility and accuracy of the findings. Once the data collection and analysis were completed, a summary of the findings (2-page summary) was emailed to each participant who will had 1 week to review the findings. The participants were instructed to check the findings for accuracy of their data. This process helped detect any unchecked bias or misunderstandings (Merriam, 2009). Internal validity and credibility of the study was strengthened using member checking (Merriam, 2009).

Data Analysis

I analyzed the data thematically by employing *a priori*, open coding, and axial coding strategies. The data for coding the audio recordings of the interviews was transcribed by me using Dragon Speech software. The use of a NVivo 12 QDAS software was used to help organize and store the data. The transcripts and documents were imported into NVivo 12 for data analysis.

The first cycle of coding the data included the *a priori* and open coding methods. As part of initial coding a list of codes (*a priori* codes) were created beforehand from

TALiP framework to align the data with the framework and RQs. The open coding method was used, concepts and categories were created based on the data from the interviews and documents.

The second cycle of reading the data used the axial coding method. This method uses the first cycle codes to find the most dominant categories in the data (Saldaña, 2013). This process helped to further define the characteristics and scope of the categories. This helped me see the relationship between the categories and subcategories and discover links in the data (Saldaña, 2013). Themes emerged from the existing codes and categories.

Evidence of Quality

Strategies for promoting validity and reliability in the study are necessary for those outside of the study to trust the study was conducted with high ethical standards. Decisions around the quality of data and the analysis of the data can help the trustworthiness of the findings of the study based on what strategies are used by the researcher (Merriam, 2009).

The level of reliability of the study indicates the consistency of the findings. One strategy to increase the reliability of a study is data triangulation where the data from the interviews were corroborated with the data from the document review (Patton, 2002). Strong evidence appeared in both sets of data, interviews and document review. This form of triangulation will also reduce vulnerability in the data and will test for consistency in the data which will provide a more in depth look in the relationship between the data and the problem (Patton, 2002, p. 248).

Treatment of Discrepant Cases

Discrepant cases are those which data do not fit with the other data, either collected by the researcher or from other studies (Patton, 2002). Patton asserted these types of cases can help guide the analysis of the data by confirming the findings and by placing boundaries around the findings (2002). These cases can help redefine the themes and results and make the study overall more robust. Discrepant cases were reported in the findings; specifically, how the cases do not fit with the other cases and what the data might indicate in the overall findings.

Data Analysis Results

In this subsection, I describe the context of the local problem. The data for this research study were gathered from the semistructured interviews and assessment examples documents from the 12 middle school music educator participants. Data were analyzed thematically analysis using a priori, open, and axial coding strategies to derive themes. I describe the details of coding strategies along with how the RQs were answered in the context of the themes.

The local problem consisted of middle school music educators having a gap in AL knowledge and skills. These 12 educators were interviewed to discover the type of assessments used and what their assessment practices are in their classrooms. I transcribed the interview data verbatim, reviewed the transcripts for accuracy, and saved into NVivo 12 software to store and manage data. The a priori coding strategy was the first cycle of data analysis. A total of seven a priori codes were selected from The Knowledge Base and Teacher AL in Practice components from the TALiP framework

and words and/or phrases were highlighted based on those codes. A sample of the a priori codes, participants, and excerpts are contained in Appendix E. This process was completed throughout all the interview data.

Open coding was the next cycle of data analysis. The a priori codes and raw data were read to find repetition of ideas, terms, or phrases. Twenty-one open codes were discovered, and the data were highlighted corresponding with each code. A complete list of the open codes and corresponding excerpts and participants are included in Appendix F. Following the open coding cycle, axial coding strategy was used to analyze the data. Axial coding strategy focused on finding relationships and similarities among the open codes, and the excerpts and four codes were derived from this cycle (see Appendix G). Upon examining the axial categories, I identified two themes (see Appendix H). The themes that emerged are (a) middle school music educators use formative and summative assessments to assess student performances while aligning the assessments' purpose to type of assessment and grading practices and (b) middle school music educators use knowledge of assessment grading and student involvement feedback practices to improve student performance.

Theme 1

Alignment of Assessment Purpose to Type of Assessment

The first RQ is what are middle school music educators' perspectives of their assessment knowledge? The first theme provides insight into how this group of educators used their knowledge of assessment (formative and summative) to better student performance and concert preparation. The music educators used formative assessment in

the classroom rehearsals to improve students' skills and performance level as well as providing student accountability. CHOIR03 shared her perspective about formative assessment during classroom instruction,

...[music educators assess] every day, because you have to be listening. . .

if you aren't listening the kids are going to be singing or playing the wrong

notes. . . . you're constantly engaged. You're constantly assessing the kids.

CHOIR03's point is that the purpose of the constant use of formative assessment during classroom instruction is to correct students' performance skills and make students accountable for their part in the group performance. During this period of instructional time, music educators used this type of assessment to guide the students toward a higher performance level through the assessment and feedback process. The constant assessment in the music classroom fits the purpose of performance improvement.

The middle school music educators used their knowledge of playing or singing assessments to align with their knowledge of music performance standards and used these assessments to hold students accountable for their learning. BAND01 reported how "[playing assessments] allows me to hear every kid play and keep tabs on how they're doing. Because in a large group it's impossible." This is reaffirmed by ORCH04 who stated, "when I start [listening to] students, especially when they're doing tests, I can see [the students] don't understand this [musical concept]." Play and singing assessments are used by the music educators because they believe this is the best way to assess students' knowledge and performance of musical skills and continue to hold them accountable for their individual performance. Based on the findings, I concluded that multiple middle

school music educators throughout their careers increased their frequency of playing assessments to monitor student growth and progress. Both BAND04 and ORCH03 shared their belief that increasing the frequency of assessments was more beneficial, as stated by BAND04 who “changed to weekly [playing] tests a few years ago and just made a tremendous difference,” and reinforced by ORCH03 “[over time] I’ve [changed my practice] into very quick assessments more often.” These educators found that quick, simple, and frequent assessments provided positive student performance outcomes.

In addition to accountability and increased frequency, middle school music educators used playing or singing assessments to align with the desired outcome. The main output for music educators is the performances for their performing groups. The “concert[s] are the most important thing” (ORCH02) to these teachers. The view of these educators is that “what [students] can do on a paper doesn't really necessarily transmit to what we actually do, . . . there has to be playing. That's our purpose” (BAND04). Students must demonstrate their abilities through the same medium which the performance will occur. This alignment of performance type and playing or singing assessments is important to accurately prepare students for performances. BAND03 mentioned how close to performances “all of [the] assessments are playing checks of [the performance] music” to help prepare the students for the performance. This way the music educators can make corrections, reinforce good habits, and ensure the students are prepared in time for the performance. These middle school music educators assessed students using both formative and summative playing or singing assessments to increase student accountability and for concert preparation. Additionally, music educators

increased the frequency of the assessments to help students increase their musical skills and abilities.

Eleven out of the 12 participants interviewed had developed a routine of how often assessments would occur in class. Of these, most created a weekly assessments schedule, as mentioned previously by BAND04 and ORCH03 who have weekly playing assessments. Other educators also had a schedule of having weekly assessments. CHOIR01, CHOIR02, CHOIR04, ORCH02, ORCH04, and BAND03 all have weekly assessments with BAND03 and CHOIR04 who developed an assessment or task assigned for each day of the week. The purpose for these routines and frequency is used to assist the students' growth in musical skills and prepare them for performances (BAND03 & BAND04). A common practice among these music educators is to have weekly participation points which are used to assess students' preparation and attitude. There was no standard definition among the participants of how this type of assessment was graded. Participants used a series of yes/no questions to assess participation and award participation points. For Example, "did you do your job, were you prepared, [did] you [have] your instrument?" (CHOIR02), "Do [you] come to class? Do [you] participate? Are [you] responsible? [Did you] bring [your] instrument? Do [you] behave?" (ORCH03), "[do you] have pencils on [your] stand, shoulder rests for the upper strings, . . . [do you have] all [your] music?" ORCH01, and "[Did you] show up with [your] instrument prepared to play and . . . actually engag[ed]?" (BAND04). Through the routine of weekly assessments, including participation points, these music educators

aligned their purpose of regularly assessing students to foster growth and concert preparation.

Alignment of Assessments to Test Purpose and Grading Practices

At the center of the grading of assessments are the criteria which the music educators use to grade the students' assessments. The ultimate outcome of instruction is that students "master . . . the skill[s]" (BAND02). Three of the participants emphasize how they "make it so [students] can always retake a test" (BAND02) which places a priority on the students mastering the musical skills. This emphasis from the educator places the purpose of the assessment at the center of what the educator's priorities are, "I put the importance [or priority] on what I think is important" (CHOIR02) and the criteria by which the assessments are graded. One music educator phrased it this way, "I plan [assessments using] the backwards assessment [model]. . . I know where I need them to go for next year, and so there are benchmarks [I want them to attain throughout the school year]" (ORCH01). The purpose of assessments come from what the music educators prioritize and in turn it becomes the criteria used to grade the assessments they give.

Grading areas is another practice of the middle school music educators where the purpose comes from the educators' priorities. Grading areas are the types of assessments (formative or summative assessments) that are developed and graded. Based on the data, there are two groups of educators: those whose grading areas are aligned together within their department or school and those who are more independent and not aligned within their school or department. Those who are aligned within the school or department

worked to have the same gradings areas (playing or singing assessments, formative and summative, grading practices for performances, and so forth.). Those educators who were not aligned within the school or department did not make an effort or felt it was not necessary to align assessment types or other aspects of grading areas. BAND01 and CHOIR03 teach at the same school where alignment of grading areas was important to them. Their grade books include only summative assessments with written knowledge assessments assigned 5-to-49-point value. All playing assessments were assigned a 50-point value. They also make their winter and spring concerts the semester exam which was 10% of their semester grade (the two quarters for the semester were worth 45% each of the semester grade). All these decisions were made collaboratively among educators. BAND03, CHOIR04, and ORCH03 teach at the same school and aligned their grading areas into four major categories: (a) practice logs, (b) participation, (c) playing tests, and (d) classroom music knowledge worksheets (BAND03). These educators also received special permission from their administration to change the formative and summative ratio from the school mandated 80% summative and 20% formative to 60% summative and 40% formative. These music educators wanted to place the emphasis on the practicing and the process of learning which was contrary to the first mentioned group who emphasized the performance.

The other middle school music educators did not place any importance on being unified with their music colleagues. ORCH01 and BAND02 both teach at the same school but did not choose to unify their grading areas. ORCH02 and CHOIR02 also teach at the same school and had differing grading areas and ratios of formative and summative

assessments. The other three educators did not have colleagues from their school participate in this study, but through the interview process did not share they unified their grading areas with other music educators at their school.

This first theme is about the perspectives of these 12 middle school music educators of and how they used their expertise and knowledge to shape their assessment choices. They used their knowledge of assessment to align with the assessment's purpose to increase the performance level and skills of their students. These music educators used playing or singing assessments, increased the frequency of assessments, and had a focus on the public performances to guide the alignment of the assessments. The purpose of the assessments was guided by the criteria which was set by educators' music priorities. Grading areas was another piece of evidence by which the priorities of these music educators were noticeable. All these data points make discernible the perception of the participants and provided insight into research question number one.

Theme 2

The second RQ features how middle school music educators use assessments to evaluate students' knowledge and skills. These middle school music educators use their knowledge of assessment grading and student involvement in feedback to improve student performance. The grading practices of these educators fell within the two main areas of rubrics and student-centered philosophy. The feedback practices included immediate verbal and written feedback strategies.

Knowledge of Assessment Grading

Central to the grading practices of these middle school music educators' knowledge of grading was the use of rubrics in their grading. Eleven out of the 12 participants used rubrics when grading playing or singing assessments and would discuss with their students the rubric used at the school district sponsored yearly assessment festivals. Of the 11 who used a rubric when grading, four used the school district created festival rubric, one used a rubric developed by the publisher of a method book, and six participants used a rubric which they created. Some of the educators pointed out how they used rubrics to help with their grading prior to, during, and after assessments. Both CHOIR04 and BAND03 give a copy of the rubric to students beforehand so they knew how they would be graded. Other participants mentioned they discussed what is on the rubric with the students but did not mention giving the students a copy of the rubric. CHOIR04 also asserted, "I've found that if I'm able to really lay it out and explain [the rubric] to the [students] beforehand, they have a better understanding . . . about what I hope that they achieve." During the assessment, BAND01 uses the rubric to remove any bias she might have as a teacher. By grading using the rubric, BAND01 can remove any factors that might deter grading from the intended goal, "[I] grade them based on just the rubric and not any sort of extra stuff." After the assessment, CHOIR03 and CHOIR04 mentioned how they used rubrics to easily explain to students and parents how a particular grade was earned.

A common foundational idea discussed in each middle school music educators' interview was the philosophy that students were at the center of their educational choices,

including grading practice, a student's musical independence, and musicality. Student-centered grading practices was exemplified by BAND01,

...if a student's having trouble with something, I try and give them a way to be successful, whether it's coming in and getting help or pairing them up with an older student who can work with them, [or] giving them a chance to retake tests.

CHOIR03 and ORCH01 used formative assessments to assist students who were struggling with the content and focus on those students with “after school sections” (CHOIR03) and “work[ing] individually with [students]” (ORCH01). When grading assessments, BAND02, BAND03, and ORCH02 consider the challenges some students need to overcome in and out of the classroom as well as the growth and effort students make in class. BAND03 stated his philosophy of student-centered grading in this statement, “I leave room for subjectiveness [in my grading] . . . I know how far he's come; I know his home life, I know all the walls that are up against this kid.” These music educators wanted all students to become successful through multiple test retakes, formative assessments results, and grading which supported the students' home life.

Another way which these music educators placed students at the center of their educational philosophy is the desire to see students become a more independent musician; meaning students have a set of musical skills which allows them to learn and perform music at their level with minimal help from a teacher. These educators show this desire through their classroom policy of allowing test retakes. Test retakes focus on the student learning the musical skill and continuing to work towards the mastery of musical

skills. BAND02 allows students to retake assessments and mentions why. “[I] want the mastery of the [musical] skill . . . I'm going to make it so they can always retake a test.” CHOIR03 wants this musical independence to last throughout the students’ lives, “I think it's important for the kids to know how to read music by themselves. I don't think they need an adult to teach them how to read music their entire life.” Assessments are necessary to measure how students are achieving and progressing towards achieving an increased musical independence. Music educators achieved this objective by monitoring the students’ music knowledge and setting goals. ORCH02 and ORCH04 shared how this desire to have the students be independent required individual assessments to “gauge where [students are] individually” (ORCH04). Regular assessments, whether weekly as with BAND04 and CHOIR03 or every two weeks as with ORCH02, gives these music educators the information needed to monitor the students’ musical independence. To help students become independent in their learning, CHOIR04 uses goal setting to help students create ownership for their learning. CHOIR04 ask the students “how would you improve your performance? What personal goal can you set for your future performance and then what do you need to reach that goal?” Included with these questions are class discussions and teacher input to help the students set and achieve goals. Through individual assessments, goals, and music educators who set up their classroom for mastery of skills, these music educators focused on their students increasing their musical independence.

The goal of these music educators was for students to use musicality, how the musician uses musical concepts to connect with the listener, in their performances.

BAND04 emphasized this using the school district festival rubric and discussed the need to perform musically,

We talk about what makes it superior rating [the highest rating a group have be given] is beyond the notes and rhythm. . . . What makes [a superior performance] that way? Musical expression. . . trying to shape [musical] lines.

CHOIR03 focused on musicality by making sure the students learn musical concepts. The students learn “key signatures and vocabulary so that they can really focus on making music and musicality.” Once students understand basic musical concepts, they are better prepared to perform music in the correct style or mood. The students will also have a higher capacity for performing like a professional musician, who uses all musical concepts to connect with the audience in a musical experience. BAND01 emphasized students’ progress from learning musical concepts to performing musically to performing musicality. For example, “advanced band should be based on their musicality” since they have worked on “individual pedagogical” musical skills. The concept of musicality is a higher performance level outcome which these music educators aspire for their students. This is central to the educators’ student-centered philosophy which shapes their goals for the students.

Student Involvement Feedback Practices

Middle school music educators used immediate feedback in verbal and written form on formative and summative assessments to improve student performance.

BAND02 thought it was beneficial for the students to hear immediately what went well and where corrections need to be made.

I usually give immediate feedback . . . as soon as they're done. I don't just go one right after the other, . . . students [need] to hear immediately right after what was good or what was not, what they need to work on and focus on for the next time.

BAND02 believed giving immediate individual feedback was important. This music educator provided feedback immediately as opposed to hearing all students play individually before giving feedback. "I don't just go one right after the other, I know some schools [teachers] do that and that's really efficient use of time but I don't know that it's necessarily efficient or helpful for the students." ORCH04 and CHOIR03 has similar beliefs about immediate in class feedback, "I listen while the kids are rehearsing and then . . . we fix it right then" (CHOIR03) and "[I] verbally tell the kids how they're doing, every day" (ORCH03). Each music educator identified that immediate feedback was a positive strategy to help students perform better.

Although all the participants agreed immediate feedback was important, not all teachers used verbal or written feedback. Two participants mentioned that they did not use the written feedback method because written feedback takes a lot of time during the assessment or outside of class. Most middle school music educators have large classes which makes giving written feedback a challenge. It takes ORCH01 a couple of weeks outside of the instructional day to give meaning and specific feedback to all her students. ORCH03 switched from giving written feedback to students to giving verbal because of

the time required to provide written feedback. “Periodically I give them written feedback, not very often because it does take a lot of class time to give each kid written feedback.”

ORCH02 did take the time to give written feedback to the students, “I write comments on the rubric if they lost a point . . . I try to write quickly . . . before I move on to the next student.” Other participants who gave written feedback on summative assessments use Google Classroom. Students submit a video outside of class time to Google Classroom and the educator can write feedback for students to view the teacher’s response later. This assignment was an example when immediate feedback was not given. BAND03 and ORCH01 chose this method to create more rehearsal time with the students and give precise feedback to students. BAND03 shared examples of some of his comments on Google Classroom assessments, (a) “Great job. . . keep working hard,” (b) “come see me at . . . advisory . . . let's work on this,” and (c) “make sure you kick out your third valve slide when you're playing low D on this test.” ORCH01 gave specifics in the feedback given for students. For example, “at 17 seconds [in your video recording] your wrist was caved in [and needs to be corrected]” and treats the feedback and written correspondence with the students like a “private lesson” where students would receive one-on-one instruction.

Some of the participants went further in their feedback by providing a class discussion about the assessment after verbal feedback was given by the teacher. BAND04 discussed the assessment results with the class to help students understand the range of their performance levels. In many music programs, the performers are ranked by the students’ performance of musical skills with the highest-level performer labeled as first

chair, then second chair, third chair, and down to last chair. BAND04 would rechair, or reorder, the students after every assessment and then discuss this process with the class. As a whole class discussion, students explain “Why is this [student] stronger than the other one?” After formative assessments during class, CHOIR03 had students discuss with each other about the rate of progress. This music educator used prompts to engage students in discussions, “Was that good? . . . What was good? What was bad? How can we fix that? What do you need to do? What do you need from me [the teacher]?” CHOIR04 had students grade themselves and compare their self-assessment of their performance with the teacher’s assessment of the performance, “. . . we compare, we talk about this is what you saw . . . this is what I saw.” The addition of class discussion was a strategy these music educators believed would assist the students to increase their musical skills and performance.

The second theme, the grading and feedback practices of these middle school music educators, answered the second research question of how these educators used their knowledge of assessments for grading to assist students developing independent musical skills and musicality. The feedback practices furthered this purpose of students increasing their musical skills and performance level.

The two RQs were answered in the context of these two themes. The middle school music educators used their knowledge of assessment to better student performance through aligning the purpose of assessments to the type of assessment and grading practices along with their knowledge of grading and feedback. The participants used constant formative assessment during class rehearsals to better students’ musical skills.

The music educators used playing or singing formative and summative assessments to hold students accountable for their learning with an increased frequency to help prepare students for performances. Assessment routines were used in the classroom and rubrics were used to grade the assessments. The music educators wanted students to obtain the necessary musical skills and chose strategies to help the students increase their musical skills with a focus on musical independence and musicality. Immediate verbal and written feedback along with class discussions were strategies used to focus on the students' learning and increased musical skills and performance level.

Summary of Findings

I examined the data from the semistructured interviews and the document collection. From the data emerged the following two themes:

- Theme 1: Middle school music educators used their knowledge of assessments by aligning assessment purpose to type of assessment and aligning assessments to purpose and grading practices.
- Theme 2: Middle School music educators use knowledge of assessment grading and student involvement feedback practices to improve student performance.

Theme 1

The music educators of this study used their knowledge of assessment purposes, methods, and context during classroom instruction. Formative assessment was consistently used during rehearsals to inform both teacher and student about students' music knowledge. These educators also used their knowledge about learning goals, what

they wanted their students to know and when, to inform what type of assessment to use. Summative performances were often the driving force for educators to choose playing or singing assessments as the main choice of assessment type. Teachers needed to hear how each student was learning in preparation for the summative playing or singing performance as well as to monitor students' growth of musical skills. These music educators viewed these assessments as accountability measures. Some educators viewed regular weekly assessments to be the best practice to increase and monitor growth. In addition to these assessment strategies and types, 11 out the 12 educators developed a routine in administering assessments. Either regular weekly or bi-weekly assessments were scheduled along with daily participation points given to assess student preparedness and participation in class.

Alignment of Purpose of Assessment and Formative and Summative

Assessments. The knowledge and use of formative assessments and the type and frequency of assessment used were supported by this study's framework. Teachers who possess AL know which type of assessment to select, the assets of each assessment type, and what the results are and how to use them (Xu & Brown, 2016). The assessment types frequently used in music education are formative and summative assessments. Teacher benchmark, diagnostic, and pretests are frequently used as formative assessments (Dixson & Worrell, 2016). Knowledge of summative assessments is used to provide teachers with students' depth of understanding and an indication of learning goals (Xu & Brown, 2016). In a music classroom the use of formative assessments occurs continuously during the rehearsal providing information to adjust the instruction and assignments for the

students. For music educators, this knowledge of formative tests can include not only students' depth of music knowledge, evidence of music learning goals, and music state standards. This understanding is supported by the knowledge base in the conceptual framework (Xu & Brown, 2016) and from current literature (Dixson & Worrell, 2016).

Based on the data from formative and summative assessments, teachers used the test results to make decisions about instruction, set students' learning goals, and provide feedback to students. This finding is supported by teacher literacy in practice in the conceptual framework. Music educators use the data from assessment results to guide classroom rehearsals to improve students' performance level (National Standards).

Alignment of Grading and Purpose of Assessments. The alignment of the assessment purposes with the grading is crucial in the teachers' AL. Teachers must not only know when to use formative and summative assessments but also what grading method should be employed to align with the purpose of the assessment (Xu & Brown, 2016). For example, music teachers want their students to be prepared for a music performance would use formative assessments to track the students' progress or assessment of learning (Black & Wiliam, 1998; Dixson & Worrell, 2016). Those assessments would be weighted in the grade book in correct proportion with the summative assessments which are based on the final music performance. Likewise, the teacher would assess the musical skills needed to perform the selected music and would use a rubric with the criteria selected from those musical skills. Most of the participants in this study were in alignment with the TALiP framework and scored formative assessments in correct relationship with summative assessments. However, there were

four participants who did not follow the correct grading practices between formative and summative assessments (Dixson & Worrell, 2016). Further professional development is needed to train a portion of the teachers correct grading practices to increase the AL of teachers.

Theme 2

Knowledge of Assessment Grading. Teachers are required to have a working knowledge of grading and feedback practices for effective AL. As described in the TALiP framework, teachers need an assessment knowledge of varying types of grading and feedback practices as well as knowledge when to use each type (Xu & Brown, 2016). Teachers who exhibit a high AL have different methods for creating assessments, such as criterion-referenced, norm-referenced, and ipsative-referenced assessments (Xu & Brown, 2016). In a music education classroom, AL would be evident in the educator administering assessments which focus on students' performing the musical selections (criterion-referenced) and then rank students based on the assessment results (norm-referenced). The educators would have knowledge of how to create assessments which focus on the various musical criteria such as intonation, balance, blend, tone quality, and so forth. Additionally, they would have knowledge on how to create assessments which focuses on one or two musical criteria and then compare students with their peers. Overall, participants' knowledge of grading was evident in this study was broadly supported by the TALiP framework; however, there are gaps in participants' AL. For example, the participants used their knowledge of grading to assess students' performance using criterion-referenced assessments as the main form for creating grades

but ipsative-referenced was not used and norm-referenced grading was only used by one participant. Knowledge of ipsative-referenced assessments would give the music educator an understanding of the rate which students believe they are comprehending the course material and in turn direct classroom instruction. The benefit of music educators knowing how to use norm-referenced assessments is to group students based on musical knowledge and skill. The educator can then help similarly grouped students with their specific needs.

Additionally, teachers need to know whether they are assessing academic or affective performances of students (Xu & Brown, 2016) regardless of the content taught. Some music educators not only assess musical performance but also grade affective performance with daily participation points. Most of the participants did grade affective performances of students as daily participation points but without following any standard of grading for these types of assessments. All educators, who used participation points in formative assessments, emphasized students' positive attitude and being prepared each day by bringing the required materials in class each day. However, researchers (Pellegrino et. al, 2015; St. Pierre & Wuttke, 2017) questioned using nonmusical assessment criteria to determine music achievement. Unless a set of standards for grading are established, assessment validity and reliability and musical achievement are questionable.

Validity and reliability are not the only AL concerns in teachers' knowledge of grading. The principles of creating rubrics and techniques of scoring assessments with objectivity and consistency are an essential component to AL knowledge in grading (Xu

& Brown, 2016). Music educators use rubrics and scoring techniques which focus on grading musical criteria, including criteria which allows for consistent grading with objectivity. Based on my study's findings, the use of rubrics was widely used among these participants, but the use of rubrics could be more effective. For example, the educators were not explicit in the definition of the criteria for each assessment, and in turn students could not make connections between the learning goals and their performance to make adjustments. Rubrics also allow for adaptation to meet student needs, such as adjusting for ability level of students, type of ensemble or style of music (Wesolowski, 2013). Again, the participants used the rubrics in a less effective method using the same rubric for all students in all settings. Another advantage for music educators to use rubrics is the ability to grade the complexities of musical performance in a way that gives meaningful feedback to students (DeLuca & Bolden, 2014; Wesolowski, 2013). Even so, the participants mostly focused their assessing on the less complex components of musical performances, such as correct notes, correct rhythm, and tone leaving out the more complex components dynamics, intonation, balance, and blend. Rubrics are designed for the educator to choose what criteria to assess (Wesolowski, 2012); however, in my study most of the participants never changed the criteria. This is where the educator can maximize or minimize criteria based on the needs of the students through the flexibility of the rubric criteria (Richerme, 2016).

Rubrics are also a tool by which music educators can grade musical performances with objectivity and consistency (Wesolowski, 2015, 2016; Wind, 2016). Music educators set the criteria of the rubric and use that rubric to assess each student's

performance based on the predetermined criteria and thus establish practices that are objective and consistent. The participants used the rubric as a tool in this way when they assessed student performances relying on the rubric as a tool for judging how students fared in each criterion.

Knowledge of Assessment Feedback. To help students learn, teachers need to know the principles of feedback which includes the different types and when to use each type (Xu & Brown, 2016). In the setting of a music classroom, the music educator uses feedback during times of continuous formative assessment, as in large group rehearsals, after summative performance assessments, and includes feedback to a large group of students as well as individuals. Feedback comes in the forms of verbal and written form and includes immediate and differed feedback (Heitink et al., 2016). Music educators give the type of feedback which best applies to increasing student confidence and level of musical performance. The participants of this study valued immediate-verbal feedback since the teachers gave meaningful feedback within the constraints of large class sizes and limited time. This group of educators appeared to be highly skilled in giving feedback which were both constructive in the artistic process and supportive to student motivation and aligned learning with the National Arts Standard MU:Pr5.3.E, which focuses on refining musical skills for performance.

Project Deliverable

Middle school music educators use formative and summative assessments to improve student growth and achievement of musical skills. The outcomes from this study are identified as a need to fortify music educators' AL by aligning their grading practices

with the purpose of assessment, modifying how nonmusical skills are assessed, and using rubrics effectively to assess student performances. By addressing these three outcomes, middle school music educators can more fully develop their AL.

The project deliverable was a white paper to middle school music educators and school leaders who supervise them. The white paper focused on the outcomes from the study with recommendations for possible solutions. The white paper will be distributed to middle school music educators and presented at one of the annual music teacher meetings. The three recommendations are:

1. Alignment Of Grading Practices with The Assessment Purpose
2. Grading Of Nonmusical Achievement
3. Use Rubrics in Grading Assessments

The middle school music educators and school leaders can use the white paper to reflect on their own assessment practices and how they can develop more fully the AL of middle school music educators.

Conclusion

Section 2 contained the description of the research methodology, the data collection, and data analysis. A qualitative case study was used to answer the RQs using data from interviews. Two themes emerged from the data analysis and were the basis for the outcomes and the type of project for this project study. Section 3 will describe the project in detail and will include a literature review linking the type of project and the outcomes from the study.

Section 3: The Project

Introduction

The purpose of this qualitative explanatory case study was to explore middle school music educators' AL in the areas of assessment knowledge and practices based on Xu and Brown's (2016) TALiP framework. I used the findings of this study to write a white paper targeted to middle school music educators and their school leaders. The findings from this study show that there are three components that middle school music educators can strengthen to further develop their AL. These components are the alignment of grading practices with the assessment purpose, the grading of nonmusical skills, and the effective use of rubrics. These three components are the basis for the recommendations in the white paper to support the AL of middle school music educators and their school leaders. Section 3 includes the description of the white paper, the rationale for using a white paper as the project, a review of the literature relating to the recommendations, the project evaluation plan, and discussion of the implementation and implications of the project.

Description and Goals

The project deliverable for this study is a white paper targeted to school leaders and music educators and aimed at improving the AL of music educators. The purpose of the white paper is to offer recommendations to middle school administrators and music educators for implementation of literacy assessment based on the findings from this study. The goals of the paper are to further middle school music educators' understanding of how to improve their AL by aligning the type of assessment with assessment purpose,

nonmusical assessments, and use of rubrics. If music educators understand and apply the recommendations in the white paper in the classroom, they may be able to provide, through their assessments, more accurate data about student growth and achievement of musical skills.

The white paper begins with a discussion of why AL is relevant in the current educational context. To explain the relevance of AL, I describe the problem, which is the gap in middle school teachers' assessment knowledge and practice. I include a definition of AL featuring detailed descriptions of current assessment knowledge and assessment practices from scholars in the field of assessment. The three recommendations are explained with evidence from the study and are supported by scholarly literature. I provide examples for each recommendation to help music educators apply these recommendations to their current situation. I conclude the white paper with a description of educational benefits for teachers and students when educators develop their AL.

Rationale

A white paper is a position document that is written to give possible solutions or policy suggestions to a problem (Hyde, n.d.). It is written to “describe a position on an issue and the rationale for that position” (Xavier University Library, 2014, para. 1) and should have a clear purpose focused for the intended audience (Dodge, n.d.). For these reasons, I selected a white paper to communicate the recommendations to stakeholders at the research site. Using a white paper, I sequenced the contents of this project to include the problem explored by this qualitative case study, the recommendations, and the potential benefits of the recommendations.

Hyde (n.d.) stated that a white paper should include needed information that is clearly stated and solutions that are presented in an easy to navigate document. In developing this white paper, I considered Hyde's recommendations and displayed quotes and diagrams making the document visually easy to navigate. Additionally, I was aware of the district leadership's professional development on formative and summative assessments and development of AL of teachers. With this background knowledge of the necessary elements of a white paper and the professional development focus, I ensured that the contents of this project align with the assessment goals of the school district.

Review of Literature

This subsection is a scholarly literature review of white papers and the topics connected with the outcomes of my research. I conducted the literature search by using Google Scholar and databases through Walden Library. The following search terms were used: *white paper*, *white papers in education*, *assessment policy education*, *assessment policy arts education*, *rubric secondary music assessment*, *assessing music performance*, *model cornerstone assessments*, and *assessment in music education*. Other search techniques included using "the cited by" feature in Google Scholar, the recommended articles on journal websites, author searches, and keywords in the article description. The following literature review is divided into topics containing current discussions about white papers, grading alignment, grading of nonmusical achievement, and rubrics for assessing musical performances. Each of these topics is relevant to the problem studied.

White Papers

Researchers compose white papers to make policy recommendations, present

data, or propose a solution to a problem (Hyde, n.d.). They are used mostly in business, education, marketing, or technical industries (Campbell & Naidoo, 2017; Hyde, n.d., Malone & Wright, 2018). There are no standards regarding the format or content that authors must meet to define their document as a white paper (Campbell & Naidoo, 2017). However, there are common features of white papers, which include medium, distribution, length, tone, and perspective (Malone & Wright, 2018). White papers should focus on a specific problem with a clear purpose written for a specified audience (Hyde, n.d.). It should include visual elements such as charts, tables, bullets, or numbered lists and should be easy for the audience to navigate with informative headings. Most importantly, the white paper should not only present the problem with data to support it but should describe why the solution is important and effective (Dodge, n.d.).

In 2018, the National Art Education Association published a set of white papers on assessment in art education (Etheridge, 2018; Galbraith & Bobick, 2018; Gates, 2018a, 2018b; Giobbia, 2018; Guenter, 2018; Hu, 2018; Meier, 2018; Ní Bhroin, 2018; Sabol, 2018a, 2018b; Sickler-Voigt, 2018a, 2018b, 2018c; Tomhave, 2018). These white papers are examples of how white papers can be used to describe the current setting of art education. Each one includes the features discussed by Malone and Wright (2018), such as section headings and visual graphics; is a brief document; is written in an authoritative tone using PDF format; and is user focused.

Meier (2018) offered recommendations for white papers. The author first gave a brief introduction about the need for art educators to use qualitative assessment in the classroom to move away from assessments that focus on artistic structure and move

towards assessments that focus on ideas. The first page includes graphics along the top and left margin with pullout quotes to make the white paper visually appealing. The problem is presented and is followed by three recommendations and a conclusion. The document is concise and brief and has headings for each section for easy navigation. This format is similar to the format used in this white paper project.

Alignment of Grading With Assessment Purpose

Xu and Brown (2016) described knowledge of assessment purposes, content, and methods as teachers knowing why they are assessing students, whether it be a formative or summative assessment. Teachers must also have knowledge about different assessment methods and assessment strategies. One of the findings from this study was the wide variety in how middle school music educators treated formative and summative assessments. The purpose and the types of assessments given by this group of music educators did not align correctly creating misalignment in the gradebook.

Researchers McCallum and Milner (2021) used qualitative methodology to evaluate the effectiveness of formative assessments and staff reflections of the data gained from the assessments. In two 1st year university courses, McCallum and Milner used optional e-formative assessments followed up by a questionnaire about the students' perceptions of the usefulness of the e-assessments. Staff then reflected on the results of the e-assessments throughout the semester. The formative e-assessments helped the students monitor their own progress, encouraged them to study, and increased the students' perceived level of learning and understanding (McCallum & Milner, 2021). The staff benefited from the e-assessments as well. They monitored student understanding and

provided support to struggling students as well as gave feedback about the use and content of the e-assessments. One drawback from the staff perspective was the time it took to create the e-assessments. This study supports the effectiveness of formative assessments and how teachers can use the data to monitor students' growth and adjust instruction. When music teachers understand that formative assessment is used to monitor student growth and not the final learning product, they may be inclined to correctly align their grading practices to place formative assessments as scaffolding toward the pinnacle summative assessment.

For music teachers to align grades with the purpose of assessments, they not only need to understand correct methods for alignment but also some of the common mistakes made by music teachers. Denis (2018) conducted a literature review which explored assessment in the field of music education. Like McCallum and Milner (2021), Denis agreed that there is a benefit to students receiving formative assessments with feedback but cautioned that when music teachers use formative assessments, they must correctly align the assessment purpose with the grading and provide a learning action following the assessment or they are mistakenly using a summative assessment. In providing an example of potential grade misalignment, Denis warned about the danger of music teachers having "deficiencies in practitioner assessment knowledge [which] may lead to misuse of assessment strategies, and in turn promote incorrect conclusions and negatively influence instruction" (p. 26). Denis also encouraged music educators to use multiple forms of assessing musical performances to overcome the subjectivity that can accompany musical assessments. Using multiple forms of assessments in this way

triangulates the data from the assessments. Teachers can then confirm with multiple types of data the correct, or incorrect, alignment of the assessment purpose and grading practices.

Teachers use formative assessments correctly when they assess the learning of students, provide feedback, analyze assessment data, and then adjust instruction based on the data (Black & Wiliam, 1998; Dixson & Worrell, 2016; Payne et al., 2021). However, for these assessments to provide meaningful data they need to be aligned with standards. When the assessment is aligned to the content standards and the grading is aligned with the assessment purpose, the grades which students receive will be an accurate representation of the knowledge and skills of the students. Payne et al. (2019) support the need for assessments to be grounded in and aligned with content standards.

Payne et al. (2019) emphasized that assessments need to be aligned with standards and embedded throughout instruction. These assessments should be preplanned and the data should be used to document students' learning. Willerson (2018) claimed that the arts educators can align assessments to the National Core Arts Standards (NCAS). He further argued that the NCAS are a bridge between nonarts policy and the legitimacy of the arts curriculum and practices of the arts educators. For example, the NCAS contains the language of assessment policy along with the language of the artistic processes. Using this language in the NCAS documents provides legitimacy to nonarts policy makers. This language was missing in previous arts standards which brought legitimacy concerns to nonarts policy makers (Willerson, 2018). Another element of the NCAS connecting the arts standards with nonarts policy makers is the Model Cornerstone Assessments (MCA).

Payne et al. viewed the MCAs as a part of the evolution of the assessment culture of music education and reminded music educators that an “understanding of and competency with effective assessment processes ensures student learning at all levels and creates strong and dynamic music programs” (2019, p. 41). Music educators may have stronger AL if they understand the alignment NCAS with assessment.

Grading of Nonmusical Achievement

Grading of nonmusical achievement is the assessment of attendance, attitude, or perceived student effort as part of the letter grade (Myers, 2020; Pellegrino et al., 2015; St. Pierre & Wuttke, 2017). While the intention of grading these areas is worthy, this does not reflect the level which students understand and perform musical skills. This type of grading is not standards based, does not guide instruction, and may delegitimize grades if students do not need to earn their grades (Denis, 2018; Myers, 2021). Myers (2021) promoted the new NCAS to provide guidance to music educators on what skills to assess. Assessments direct student and teacher focus (Denis, 2018). If teacher assessments are not on musical skills, then students and teachers will not concentrate on the most standards, curriculum, or performance.

In earlier research, Pellegrino et al. (2015) and St. Pierre and Wuttke (2017) agreed that secondary music educators’ grading practices were not representing musical achievement by including nonmusical skills with the grading of musical skills. This type of grading masked the actual musical skills students achieved since assessments of attitude, attendance, and preparation were included with musical skills. Myers also agrees that this type of grading has been a problem in the field of secondary music education. He

emphasized in his literature review that if music educators would place an importance on musical achievement by having grades represent what musical skills were achieved, that it would help educators “in creating curricula that value skill and learning over showing up” (2021, p. 4) Some school districts, like the location of this study, have a citizenship grade which can be used to assess the nonmusical achievement categories like attitude, attendance, participation, and class preparation.

These nonmusical achievement categories are important to the success of music programs where rehearsals are the main form of classroom instruction (Myers, 2021; Orzolek, 2020) like music educators used in this study. However, there are not many resources to guide how the grading of these nonmusical categories should be conducted by music educators (Orzolek, 2020). In creating a rubric to fill this gap, Orzolek used researcher Robert Kelley’s article “In Praise of Followers” as the foundation for how to define what these nonmusical categories should look like in the classroom. Orzolek used Kelley’s idea of the followership and created the *Engaged and Effective Followership Rubric* music educators can use when assessing nonmusical categories. The rubric communicates to students and other stakeholders the definition of daily rehearsal participation (nonmusical categories) and how students will be assessed. The drawback to this type of rubric is that Orzolek (2020) does not mention how this rubric fits with the assessing of musical skills and leaves it to educators to incorporate this in their gradebooks. If music educators do have a separate grade given on report cards which evaluates students’ citizenship skills, this could be an effective method on how to calculate that grade. For this to be most effective, there needs to be a separation of

musical achievement and nonmusical achievement in the gradebook so that all stakeholders know at what level students are achieving in each category.

Use of Rubrics for Assessing Musical Performances

Rubrics were an area of weakness among the music educators interviewed in this study. Rubrics are an effective assessment tool for the reason described by Denis (2018) in his literature review of assessment in music education. Denis (2018) noted that music educators use rubrics to communicate the expectations to the students and to increase consistency in students' performance of musical skills. When teachers think about and plan assessments, teachers can design rubrics to help students know the learning goals and what will be assessed. This process can be used to help students become engaged with the material from the beginning (Denis, 2018). Payne et al. described effective assessments

...are based on four characteristics: (1) defining the expected learning, (2) determining acceptable evidence of learning by designing tasks that require students to demonstrate necessary skills and cognitive demands, (3) employing a measuring device that differentiates qualities of achievement, and (4) thoughtful analysis and evaluation of the results (2019, p. 38).

When designed correctly, rubrics can be a “measuring device that differentiates” (Payne et al., 2019, p. 38) students' achievement levels. Music teachers can utilize this type of assessment data gained from rubrics to make instructional decisions, remediate if needed, or monitor learning.

When developing a rubric for an instrumental large group performance, Edwards et al. (2019) used 20 music educators with an average of 14.2 years of teaching experience to evaluate 25 large group performances. The researchers used the four main areas from the MCAs for each music educator to rate ensemble performances using an a priori list based on research from DeCamp's study (1980) of creating a rating system for high school band performances. From their study, Edwards et al. (2019) developed a reliable and valid ensemble performance rubric. During the development of their rubric, they removed 14 of the 39 items in the a priori list because "these 14 items caused a violation of . . . Rasch measurement and do not appropriately fit the model" (Edwards et al., 2019, p. 363). Furthermore, Edwards et al. (2019) pointed out that not all criteria of the rubric are of equal value. For example, playing correct notes and correct rhythms are basic musical skills. On the other hand, correct ensemble balance and intonation are higher level musical skills and that music teachers who place an emphasis on these skills will have students who perform at the highest levels. As a result of these changes, the researchers cautioned that when using the rubric, music educators should not tally the points for the overall rating but understand that some criteria carry more weight than others (Edwards et al., 2019).

Music teachers can use the results of Edwards et al. (2019) when creating their own rubrics. Teachers can make choices about the criteria of the rubric knowing that some musical skills should be considered more weight in the grading process. Understanding that not all musical skills are of equal weight when grading will

communicate to the students, they need to prioritize the development of those musical skills.

Another method for creating rubrics as an effective assessment tool was studied by Fraile et al. (2017). These researchers concluded from their qualitative study that students better internalized the content standards and assessment criteria when they are involved in the creation of the rubric. Fraile et al. (2017) concluded that using students in the creation of rubrics could improve student self-efficacy and self-regulate learning through a discussion about the assessment criteria, standards, and expectations. This appears to be more effective than a teacher-imposed rubric where students are less involved with the creation of the rubric. This type of rubric follows the first three of characteristics of an effective assessment presented by previously mentioned Payne et al. (2019).

During the rubric creation, music educators can use assessment data from previously administered assessments to identify which goals are the most important for student learning. Myers (2021) stated that music teachers can never assess every skill during a student performance. When creating a rubric, music educators must choose and prioritize what are the most important musical skills for each assessment. Teachers can change or modify the rubric criteria on learning goals. What teachers choose to assess and what they choose not to assess communicates their educational philosophy to students, parents, and administration (Myers, 2021). They can change the rubric criteria according to the needs of the students as they rely on effectively analyze assessment data, define learning goals, and develop a set of criteria that allows for various levels of

achievement. Music educators effective use of rubrics strengthens their AL through their understanding and practices of communicating learning goals, measuring accurately students' skills, and increasing students' understanding of standards.

Project Description

Needed Resources and Existing Supports

At the time of the project, the local school district offered professional development courses and district initiatives that were focused on developing teacher AL. However, the intended audiences of these resources are traditional classroom teachers and not performing arts teachers such as middle school music educators. Within the school district current structure, secondary music is in the Humanities Division, which is in the larger umbrella of Curriculum and Instruction Division. I elicited support from the humanities director and the coordinator for secondary music, dance, and theatre on how to disseminate the white paper in the study school. I also worked with the director of research in the assessment, compliance, and monitoring division in the district coordinating with all three positions for the approval and distribution of the white paper.

The coordinator for secondary music, dance, and theatre would be the main support since this person has the most contact with school principals. This person advises school leaders in matters of secondary music topics and the hire and development of music teachers. The coordinator also directs the focus of the music teachers in secondary music in district wide trainings by selecting the topics for professional development. If it is decided to distribute a printed version of the white paper, there will be a small cost for printing and binding the white paper.

Potential Barriers and Solutions

A potential barrier to reaching a large audience for the white paper could be the minimal frequency the secondary music educators meet for professional development. Most years, there is one meeting at the beginning of the year where the secondary music educators come together to discuss the upcoming year and have professional development opportunities. Unfortunately, these meetings are not required and happen one day before teachers' contractual start day. This creates a challenge for reaching all music educators since some do not attend the meeting. Another barrier is the minimal opportunity to physically meet and discuss the white paper with all middle school music educators. This district has 51 middle schools with music programs and some schools geographically spread out. Reaching all music teachers and the school leaders who supervise them would be a challenge. While the documents could be printed and sent through the school district mail system, the lost opportunity to discuss the white paper, answer questions, and receive feedback is a barrier.

One possible solution that could work would be to contact the supervising administrator of the novice music educators of those 51 middle schools to discuss the importance of the white paper. Those school administrators could then distribute the white paper to those teachers to discuss the contents of the white paper. The process could be ongoing, cycling through the music teachers starting with the novice teachers through the most senior music teachers. Another possible solution is to distribute and discuss the white paper at the first of the year voluntary meeting and then contact those who did not attend and make a presentation to those teachers separately.

Proposal for Implementation and Timetable

The procedures for distributing the white paper in this school district is to send the summary findings and white paper to the director. Once they have reviewed the documents, I will arrange a meeting with the director of humanities and with the coordinator for secondary music, dance, and theatre to discuss a plan for distributing the white paper and answer any questions, comments, or feedback on the recommendations. During this meeting, I will present an implementation plan which will focus on distribution of the white paper to groups of music teachers starting with the novice music teachers and cycling through to the most senior music teachers. It is anticipated during the span of 4 months I would meet with 4-6 groups of educators and school leaders. At the end of the school year there will be a follow up meeting with the Director of Humanities and with the Coordinator for Secondary Music, Dance, and Theatre to report on the implementation process.

Project Evaluation Plan

A white paper was selected to describe the findings from the study and recommendations based on the findings and to share the white paper and its contents with the stakeholders. The evaluation of the project will follow school district guidelines and follow the adopted practice of formal evaluation, using a Google Form included in the project. In this school district, school district personnel are required to submit a Likert scale and short answer professional development questionnaire to rate the presentation. The questionnaire includes two response choices, disagree or agree, with short answer questions about what support teachers may need and if they plan on using the content

from the professional development in their classrooms (see Appendix A). This evaluation tool will then be used for feedback to see if teachers find the recommendations useful and intended to use some portion in the transformation of their AL.

Project Implications

Implications for Local Community

The possible social change implications of this white paper could be an increased awareness of music educators' knowledge about AL. The increased awareness may benefit music teachers by helping them to develop their AL more fully. This would create a cause-and-effect sequence where students and parents will directly benefit from teachers whose AL is more developed. These students will have a better understanding how they are progressing, learning goals, and clear communication about teacher expectations. Parents, like the students, will also better understand the learning goals and the expectations of the teacher. The cause and effect will continue as students will more accurately know where they have strengths and weaknesses. Students would progress with accurate assessments and feedback and achieve at a higher level. Parents and students would benefit from higher student achievement and performance level.

Implications for Global Community

This white paper can possibly have social change implications for the global community by helping music educators in other school districts and states to use assessment methods which more accurately assess students' musical skills. This can help those who develop curriculum to educate preservice music teachers to include assessment knowledge and practices that support K-12 student learning. This could help preservice

teachers to have a more developed AL when they become licensed teachers and in turn could benefit students to have a more accurate assessment of their musical skills.

Conclusion

The white paper project provides recommendations based on the findings of the study. In Section 3, I described the goals of the project, a review of the literature focused on the recommendations, the project and its contents, how the project would be evaluated, and the implications of the project. In Section 4, I will present the strengths and weaknesses of the project and my reflections of the project and my personal growth through the development of the project.

Section 4: Reflections and Conclusions

Project's Strengths and Limitations

All projects have strengths and weaknesses. In this section, I discuss the strengths and weaknesses of the project that I developed based on the findings from the case study that I conducted.

Strengths

The project could influence middle school music educators and help them to be more aware of their AL and how to develop their assessment knowledge and practices. The strengths of the white paper are the universality of the recommended solutions for both instrumental and vocal settings, the focus of the recommendations on the middle school level, and the ease of implementing the recommendations in a busy music educator's workload. I believe that these aspects may improve assessment in music.

The assessment recommendations of the white paper are applicable to all performing music education genres including band, choir, orchestra, jazz, mariachi, and any others that include vocal or instrumental performances. Even though there are differences between each music discipline, the recommendations from the white paper can be implemented across all performing music disciplines. The recommendations focus on applying and improving assessment techniques for the middle school level music educator. Most researchers have combined middle school and high school music educators (e.g., Myers, 2021). Although there are similarities between the two levels, there are unique problems in a middle school music program, namely the large class sizes and the maturity of the students. In this white paper, I focus on the assessments needs

specific to middle school music educators with large class sizes to support them in their unique teaching situation.

These assessment recommendations can be implemented immediately in two ways. Music educators can use the recommendations to support their AL growth, and school leaders can use the recommendations to help with school-level professional development. For the last 6 years, this school district leadership has focused on assessment practices and provided school leaders with resources for teacher professional development at the school level. The school district, however, has not provided professional development that supports or is applicable to music educators. School leaders are not given resources on how to assist music educators' AL. In my research, I found some gaps in assessment knowledge and practice and identified support needed by teachers to develop their AL more fully. The recommendations included in the white paper may help school leaders to understand what assessment practices music educators can strengthen and how to support them in doing so.

Limitations

One limitation of the white paper is the assumption that the music educator is moderately assessment literate (i.e., possesses knowledge of assessment and how it shapes instruction). For example, one recommendation from the white paper is the more effective use of rubrics. This recommendation is for music educators who already understand rubrics, how to create them, and how to use them in grading and feedback. If the educator is not familiar with rubrics and how to grade playing assessments, then different strategies would be required before the use of the recommendations in the white

paper.

Another limitation is the type of project. A white paper requires the reader to be intrinsically motivated to take the initiative to read the white paper on their own. Another project might reach a wider audience through a presentation or professional development. However, the receptiveness of the recommendations in the white paper may not be as high in a large group meeting as those who read the white paper on their own. One way to overcome this limitation could be to use the white paper as a document employed in a schoolbook club with music educators and administrators. In this way, this would give time for discussion about assessment practices and how to implement the recommendations in the white paper.

Recommendations for Alternative Approaches

In this study, I focused on middle school music educators' assessment knowledge and practices. I could have conducted a quantitative study and had the participants complete a survey about their assessment knowledge and assessment practices. Those data would be analyzed to derive AL trends among the music educators or the factors that increase music teachers' AL. Data collected from this type of study could be used to identify the most common assessments used and which are used least. The results could then be used to help music educators to diversify their assessment practices to develop their AL more fully. Another method that could have been used was including student, administration, or parent perspectives of the classroom assessments. The data could be used to compare the various stakeholders' perspectives and then identify common understandings and misunderstandings of the purposes of assessment. The results could

be used to provide recommendations on needed communication between each stakeholder and the AL needs of school leaders and music educators.

Scholarship, Project Development, and Leadership and Change

Scholarship

Prior to this project creation, I had only heard about white papers, and I had never seen or read one. In my research about white papers, I learned that they are documents that present a problem with a proposed policy or solution to a problem (Hyde, n.d.). White papers need to clearly define the scope and purpose focused on the intended audience (Hyde, n.d.). The author of this type of document should be objective and use language that avoids bias (e.g., gender-neutral language). White papers should also have a visual appeal using the data from studies to support the position of the author.

My skills as a writer were challenged in writing a white paper. I learned how to be more positive and objective about the problem and recommended solutions. My writing skills were developed further through learning how to write a position paper to the intended audience using the findings to support my recommendations. Preparing this project also helped me to think how I should present the problem and proposed solutions to help others understand the problem and recommendations clearly. I learned that I enjoy writing position papers and learning skills to make the document clear and understandable.

Project Development

To prepare for the development of the white paper, I needed to understand the role of white papers and the format for these types of documents. I learned that white

papers are persuasive and are used to describe a problem, are directed at specific stakeholders, and include recommendations to solve the problem (Hyde, n.d.; Xavier University Library, 2014). Specific to my project, I needed to include the findings from the study for recommendations on how to support and strengthen music educators' assessment practices. In preparing the paper, I needed to have a logical flow to support of proposed solutions (Dodge, n.d.). Importantly, in the white paper, I needed to focus the argument on the important stakeholders and demonstrate how supporting these recommendations is applicable to them.

Leadership and Change

The EdD program has been a transformation for me going from a consumer of research to a researcher. It has helped me to weave research into my current assessment practices. In my own teaching practices, I have added or removed practices that have a high effect on student learning. For example, using rubrics more effectively and removing the practice of participation points are two changes I have made. Because of my current research and understanding, I am confident to speak out in school settings when I see something not in line with AL research. The line from "Amazing Grace" "I was blind but now I see" represents how I feel about the change that I underwent while completing this degree. I understand areas of strengths and weaknesses in my teaching and AL. I have volunteered on school district committees and projects to use my insights and leadership skills that I have gained during this degree program.

My leadership style has changed during this time from passive to active, and I am now willing to share my opinions. I have actively helped less experienced music

educators and student teachers by being a mentor to individual teachers. Among my colleagues I have shared rubrics and assessment strategies. I was a committee member who developed the district curriculum pacing guides for newer teachers. I am more vocal in staff meetings and other trainings offering my insights and knowledge I have gained from this degree program. Prior to this experience, I would not have been as active as I am now in sharing and supporting others. I was not confident in my skills and knowledge about teaching, but this degree program has changed how I view myself, and I now realize that I have good strategies to share with my colleagues.

Self-Analysis

My personal learning during the process of becoming a researcher was drastic. When working on my bachelor's degree, I focused on the pedagogy of teaching brass, woodwind, percussion, voice, strings, and guitar classes. My master's degree in instrumental conducting focused on becoming an artist and refining my musical skills. This research degree required large leaps in knowledge and learning to investigate and think like a researcher. During the coursework for this degree, I was exposed to new ideas and concepts about different research philosophies, how to develop a problem statement, and RQs. The most helpful for me in my development to think like a researcher was reading research articles and noticing how researchers aligned the various components in their research. This skill has helped me not only as a researcher but as a consumer of research. I notice when someone's argument is strong because their logic is aligned and when someone's argument is weak because it is not aligned.

Analysis of Self as a Scholar

Throughout this EdD program at Walden, my identity as a scholar has changed to understand ideas and concepts in a more nuanced way. I can see multiple ways of analyzing a problem or solving a problem. This was evident during distant learning due to COVID-19 restrictions and having to figure out how to teach band in a distant learning model. I coupled different methods that were successful for my students and me. I have gleaned insights to how good researchers think and problem solve. Through reading journal articles and other scholarly documents, and now including my own research, I take part in general academic discussions and more specialized ones on AL. The coursework and project study have given me insights into how I can be a leader and better educator in my home school and school district. Because of the project study, I am viewed as a leader among my colleagues in my school and district. Through the development of this project, I have increased my skills of synthesizing data into a digestible format for those who need information quickly and concisely. Overall, this degree created a large positive shift in how I think about my professional skills, how I prioritize data, and how I problem solve.

Analysis of Self as a Practitioner

The process of conducting this project study has had a profound influence on my life as an educator. The reading of research articles and research methodology books assisted me in learning how to become a better educator by helping me see and understand what practices are most effective. Reading the articles has helped me to sequence what I teach in a more logical manner and focus on helping my students to

become independent learners. The AL articles helped me to reflect as an educator and investigate what my assessment knowledge and practices were and compared them with the literature. At times I changed my assessment practices as I gained a better understanding of current assessment practices from reading the literature on AL.

Currently my school district is placing a priority on AL for teachers. This has provided me opportunities in my school to share what I have learned about AL with my colleagues. My willingness to look at my practices as an educator and reflect on my assessment skills and modify them with researched practices make my quality of teaching more effective for student learning. Through this reflection I have started to use rubrics to grade student performance assessment and use formative assessments correctly in the classroom. I now have a solo night as the final assessment of the year where my students learn and perform a solo to help them become a more independent musician. I have also changed my grading practices by removing any nonmusical assessments from the grade book and having those nonmusical assessments be part of the student's citizenship grade. These changes have made me a better educator by helping my students learn more effectively.

Analysis of Self as a Project Developer

As a project developer, I learned that I have passion for helping other educators. I want other music educators to improve their assessment skills and understand AL to help them in their classrooms. I have learned about assessment knowledge and practices which supported me in the decisions I made as an educator, and I want other educators to have the same experience and support them in making better decisions. I also learned as a

project developer that I enjoy writing as a form of communicating in a position paper. It is challenging and I find that I need a lot of refinement in my writing skills but writing helps me to formulate my ideas in a concrete way to help others understand my reasons for certain position.

Reflection on the Importance of the Work

The findings of this research project have been meaningful to me by giving me an outsider's perspective on the assessment knowledge and practices of middle school music educators. This perspective has helped me to reflect on my personal assessment practices and adjust how I use assessments in the classroom. I hope that my research and my white paper help music educators in a similar way and I hope they are willing to use my ideas in their own assessment practices. The work of teacher AL is important since it is integral for instruction adjustments and student achievement. Once assessment illiteracy is changed to AL, teachers will have a larger positive influence on students learning and achievement.

Implications, Applications, and Directions for Future Research

The area of social change this project can influence is on the skills and practices of music teachers. One teacher can influence many students over the course of time. If teachers are concentrating to better their craft of teaching, their influence over students can become more positive. Teachers who improve their craft have more knowledge and skills on how to help students learn, help students to overcome obstacles, and help students' self-efficacy in learning new material. Teachers who work to improve their AL through the recommendations in the white paper may increase their influence on student

learning and in turn may positively affect students in other areas of their learning. By helping one music teacher better their teaching craft, they can influence many students with positive learning experiences.

All middle school band, choir, or orchestra teachers can use the white paper to reflect on their own practices and improve their own assessment practices. Whether the school is urban or rural, the school demographics, the school rating, or the school socioeconomical categories, the recommendations can be used by music education to better their assessment practices. The recommendations in the white paper are universal to all middle school music educators to help improve their AL. When educators are assessment literate, they are more reflective in their practices and focus on the development of their AL which supports all students' learning. This white paper has the potential to start the discussion among music educators on how to be assessment literate and help music educators to create practices that are equitable to all students. All students would receive instruction based on their learning needs. Teachers who are assessment literate will accurately assess the needs of students, interpret the data correctly, and adjust instruction based on those assessment results. Grading will be equitable and based solely on the students' acquired knowledge, skills, and performance level. Teachers may use rubrics to communicate clear learning expectations and give feedback that focuses on the standards and learning goals for each student. This will help all students feel they have the potential for improvement and success through their efforts and the support of educators.

School leaders can also use the white paper to gain insight on how to support

music educators' assessment practices and know what support to give them. Music educators are using many meaningful assessment practices in the classroom but can be supported in growing their AL in a few areas. School leaders can support music educators with professional development or informal feedback focused on grading practices and assessment purposes being aligned, the grading on nonmusical assessments, and using rubrics more effectively. These are areas of emphasis school leaders can use in discussions with music educators to evaluate AL and target areas of weakness.

Possible areas where further research of music educators' AL could include research about the most recent music standards and the assessments linked with those standards. For example, researchers may investigate the effect of the National Arts Core Standards Model Cornerstone Assessments have on student performance. These assessments were created by a national committee of educators to help music educators of all levels have example assessments and student examples to assist music educators in assessment reliability. Researchers could use a program evaluation of the MCAs to find if the example assessments and student examples are giving the music teachers the help that was intended.

Another area for further research would be how the various uses of student involvement in assessments and what effect, if any, does involvement have on student performance. Researchers could possibly use a quantitative study with groups of students. Some groups would use assessments which involved students in the assessments process while other groups would not use student involvement in the assessments.

A third area could be educators' conception of assessment and how that

influences the decisions and actions music educators make in assessments. Through interviews and class observations, researchers can use a qualitative study to inquire about teachers' views of learning, their views about affective dimensions of learning, and their views about cognitive dimensions of learning influence teachers' assessment choices.

Conclusion

Assessment literate music educators have foundational assessment knowledge and demonstrate assessment practices where decisions and action taking are used to help students learn. The students will have a better sense of their learning goals, their strengths and weaknesses, and progress. Assessment is at the center of student learning and achievement. The recommendations included in the white paper, if followed, have the possibility to make this type of scenario a reality. Middle school music educators, who are assessment literate, can develop a learning environment where student learning is the center of decisions and teachers' actions in assessment provides direction for improvement of musical skills.

References

- Aguilar, C. E., & Richerme, L. K. (2014). A descriptive study of music teacher educators' beliefs about policy. *Journal of Music Teacher Education*, 25(2), 37-49. <https://doi.org/10.1177/1057083714553986>
- Allsup, R. E. (2015). Music teacher quality and the problem of routine expertise. *Philosophy of Music Education Review*, 23(1), 5-24. <https://doi.org/10.2979/philmusieducrevi.23.1.5>
- Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability*, 26(2), 153-176. <https://doi.org/10.1007/s11092-013-9188-4>
- Atjonen, P. (2014). Teachers' views of their assessment practice. *The Curriculum Journal*, 25(2), 238-259. <https://doi.org/10.1080/09585176.2013.874952>
- Baas, D., Castelijns, J., Vermeulen, M., Martens, R., & Segers, M. (2015). The relation between assessment for learning and elementary students' cognitive and metacognitive strategy use. *The British Psychological Society*, 85(1), 33-46. <https://doi.org/10.1111/bjep.12058>
- Baas, D., Vermeulen, M., Castelijns, J., Martens, R. & Segers, M. (2020). Portfolios as a tool for AfL and student motivation: are they related? *Assessment in Education: Principles, Policy & Practice*, 27(4), 444-462. <https://doi.org/10.1080/0969594X.2019.1653824>
- Basu, O. N., Dirsmith, M. W., & Gupta, P. P. (1999). The coupling of the symbolic and

the technical in an institutionalized context: The negotiated order of the GAO's audit reporting process. *American Sociological Review*, 506-526.

<https://doi.org/10.2307/2657253>

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.

<https://doi.org/10.1080/0969595980050102>

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575.

<https://doi.org/10.1080/0969594X.2018.1441807>

Blom, D., Rowley, J., Bennett, D., Hitchcock, M., & Dunbar-Hall, P. (2014). Knowledge sharing: Exploring institutional policy and educator practice through ePortfolios in music and writing. *Electronic Journal of e-Learning*, 12(2), 138-148.

Bogdan, R. C., & Bilken, S. K. (2007). *Qualitative research for education: An introduction to the theories and methods*. Pearson Education.

Boud, D., Dawson, P., Bearman, M., Bennett, S., Joughin, G., & Molloy, E. (2018).

Reframing assessment research: Through a practice perspective. *Studies in Higher Education*, 43(7), 1107-1118. <https://doi.org/10.1080/03075079.2016.1202913>

Bunch, E. H. (1998). Quality of life in people with advanced HIV/AIDS in Norway.

Western Journal of Nursing Research, 20(4), 465-477.

<https://doi.org/10.1177%2F019394599802000405>

Burner, T. (2014). The potential formative benefits of portfolio assessment in second and foreign language writing contexts: A review of the literature. *Studies in*

Educational Evaluation, 43, 139-149.

<https://doi.org/10.1016/j.stueduc.2014.03.002>

Cangro, R. (2015). Student collaboration and standards-based music learning: A literature review. *Update: Applications of Research in Music Education*, 34(3), 63-68.

<https://doi.org/10.1177/8755123314568794>

Chen, L. K., Kim, Y. S., Moon, P., & Merriam, S. B. (2008). A review and critique of the portrayal of older adult learners in adult education journals, 1980-2006. *Adult education quarterly*, 59(1), 3-21. <https://doi.org/10.1177/0741713608325169>

Christoforidou, M. & Kyriakides, L. (2021). Developing teacher assessment skills: The impact of the dynamic approach to teacher professional development. *Studies in Educational Evaluation*, 70, 101051.

<https://doi.org/10.1016/j.stueduc.2021.101051>

Christoforidou, M., Kyriakides, L., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation*, 40, 1-11. <https://doi.org/10.1016/j.stueduc.2013.11.006>

Common Core State Standards Initiative. (2019). *Development process*.

<http://www.corestandards.org/about-the-standards/development-process/>

Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134-144.

<https://doi.org/10.1016/j.tate.2017.12.010>

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating*

quantitative and qualitative research. Edwards Brothers.

- Crochet, L. S., & Green, S. K. (2012). Examining progress across time with practical assessments in ensemble settings. *Music Educators Journal*, 98(3), 49-54.
<https://doi.org/10.1177/0027432111435276>
- Dann, R. (2019). Feedback as a relational concept in the classroom. *Curriculum Journal*, 30(4), 352–374. <https://doi.org/10.1080/09585176.2019.1636839>
- Dell, C. (2014). Musical home environment, family background, and parenting style on success in school music and in school. *Contributions to Music Education*, 40, 71-89.
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education. *Journal of Teacher Education*, 64(4), 356-372. <https://doi.org/10.1177/0022487113488144>
- DeLuca, C., & Bolden, B. (2014). Music performance assessment. *Music Educators Journal*, 101(1), 70-76. <https://doi.org/10.1177/0027432114540336>
- DeLuca, C., Coombs, A., LaPointe-McEwan, A. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation* 61, 159-169.
<https://doi.org/10.1016/j.stueduc.2019.03.012>
- DeLuca, C., & Hughes, S. (2014). Assessment in early primary education: An empirical study of five school contexts. *Journal of Research in Childhood Education*, 28(4), 441-460. <https://doi.org/10.1080/02568543.2014.944722>
- DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy & Practice*, 24(2),

121-126. <https://doi.org/10.1080/0969594X.2017.1297010>

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248-266.

<https://doi.org/10.1080/10627197.2016.1236677>

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy:

A review of international standards and measures. *Educational Assessment*,

Evaluation and Accountability, 28(3), 251-272. [https://doi.org/10.1007/s11092-](https://doi.org/10.1007/s11092-015-9233-6)

[015-9233-6](https://doi.org/10.1007/s11092-015-9233-6)

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016).

Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 25(4), 355-375.

<https://doi.org/10.1080/0969594X.2016.1244514>

Denis, J. M. (2018). Assessment in Music: A Practitioner Introduction to Assessing

Students. *Update: Applications of Research in Music Education*, 36(3), 20–28.

<https://doi.org/10.1177/8755123317741489>

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the

classroom. *Theory Into Practice*, 55(2), 153-159.

<https://doi.org/10.1080/00405841.2016.1148989>

Dodge, A. (N.D.). 8 Essential Elements of an Effective White Paper.

<https://www.copyright.com/blog/8-essential-elements-of-an-effective-white-paper/>

- Doğan, C. D., & Uluman, M. (2017). A comparison of rubrics and graded category rating scales with various methods regarding raters' reliability. *Educational Sciences: Theory & Practice, 17*(2), 631-651. <https://doi.org/10.12738/estp.2017.2.0321>
- Droe, K. L. (2014). Investigating parent and teacher perceptions of school, family, and community connectedness. *Contributions to Music Education, 40*, 57-70. <http://www.jstor.org/stable/24711071>
- Edwards, A. S., Edwards, K. E., & Wesolowski, B. C. (2019). The psychometric evaluation of a wind band performance rubric using the Multifaceted Rasch Partial Credit Measurement Model. *Research Studies in Music Education, 41*(3), 343–367. <https://doi.org/1177/1321103X18773103>
- Etheridge, J. (2018). Experimenting with Assessment: A Yearlong Process. Section I: Assessment in Art Education--Building Knowledge. Assessment White Papers for Art Education. *National Art Education Association*.
- Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation, 53*, 69-76. <https://doi.org/10.1016/j.stueduc.2017.03.003>
- Galbraith, A., & Bobick, B. (2019). A Snapshot of Portfolio Assessment. Section IV: Analyzing, Interpreting, and Reporting Art Education Assessments. Assessment White Papers for Art Education. *National Art Education Association*.
- Gates, L. (2018a). A Participatory Model for Assessing Teacher Candidate Dispositions. Section IV: Analyzing, Interpreting, and Reporting Art Education Assessments.

- Assessment White Papers for Art Education. *National Art Education Association*.
- Gates, L. (2018b). Aligning Learning Objectives and Assessment Methods in the Visual Arts. Section III: Planning and Implementing Visual Arts Assessments. Assessment White Papers for Art Education. *National Art Education Association*.
- Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2017). Assessment quality in tertiary education: An integrative literature review. *Studies in Educational Evaluation*, 55, 94-116.
<https://doi.org/10.1016/j.stueduc.2017.08.001>
- Gilbert, A. D. (2016). The Framework for 21st Century Learning: A first-rate foundation for music education assessment and teacher evaluation. *Arts Education Policy Review*, 117(1), 13-18. <https://doi.org/10.1080/10632913.2014.966285>
- Giobbia, T. (2018). Using Blogs and ePortfolios to Assess Student Growth in Middle and Secondary School Art Classrooms in West Virginia: A Cross-Case Analysis. Section II: Assessments That Promote Vibrant Learning Communities and Advocate for the Visual Arts. Assessment White Papers for Art Education. *National Art Education Association*.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues & Practice*, 33(2), 14-18.
<https://doi.org/10.1111/emip.12030>
- Guenter, C. (2019). Components for Gathering and Using Assessment Data to Inform Instruction. Section IV: Analyzing, Interpreting, and Reporting Art Education Assessments. Assessment White Papers for Art Education. *National Art*

Education Association.

- Guhn, M., Emerson, S. D., & Gouzouasis, P. (2020). A population-level analysis of associations between school music participation and academic achievement. *Journal of Educational Psychology, 112*(2), 308–328.
<https://doi.org/10.1037/edu0000376>
- Gulikers, J. T. M., Biemans, H. J. A., Wesselink, R., & van der Wel, M. (2013). Aligning formative and summative assessments: A collaborative action research challenging teacher conceptions. *Studies in Educational Evaluation, 39*(2), 116-124. <https://doi.org/10.1016/j.stueduc.2013.03.001>
- Gynnild, V. (2015). Assessing vocal performances using analytical assessment: A case study. *Music Education Research, 18*(2), 224-238.
<https://doi.org/10.1080/14613808.2015.1025732>
- Harrison, S. D., Lebler, D., Carey, G., Hitchcock, M., & O'Bryan, J. (2013). Making music or gaining grades? Assessment practices in tertiary music ensembles. *British Journal of Music Education, 30*(01), 27-42.
<https://doi.org/10.1017/s0265051712000253>
- Hash, P. M. (2013). Large-group contest ratings and music teacher evaluation: Issues and recommendations. *Arts Education Policy Review, 114*(4), 163-169.
<https://doi.org/10.1080/10632913.2013.826035>
- Hatch, E. (2020). Going Beyond “It Sounds Good:” Developing Student Capacity to Give Meaningful Feedback. *General Music Today, 33*(3), 29-35.
<https://doi.org/10.1177/1048371319892241>

- Hayward, L. (2014). Special issue on assessment for learning. *The Curriculum Journal*, 25(4), 465-469. <https://doi.org/10.1080/09585176.2014.981381>
- Hayward, L. (2015). Assessment is learning: The preposition vanishes. *Assessment in Education: Principles, Policy & Practice*, 22(1), 27-43. <https://doi.org/10.1080/0969594X.2014.984656>
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50-62. <https://doi.org/10.1016/j.edurev.2015.12.002>
- Hewitt, M. P. (2015). Self-efficacy, self-evaluation, and music performance of secondary-level band students. *Journal of Research in Music Education*, 63(3), 298-313. <https://doi.org/10.1177/0022429415595611>
- Hu, J. (2020). Poetry as Embodied Self-Assessment for Visual Art Learning. Assessment White Papers for Art Education. Section V: Visual Arts Assessments--Case Studies from the Classroom and Beyond. Assessment White Papers for Art Education. *National Art Education Association*.
- Hyde, C. (N.D.). White paper guide. https://owl.purdue.edu/owl/subject_specific_writing/professional_technical_writing/white_papers/white_paper_ppt.html
- Iusca, D. (2014). The effect of evaluation strategy and music performance presentation format on score variability of music students' performance assessment. *Procedia - Social and Behavioral Sciences*, 127, 119-123.

<https://doi.org/10.1016/j.sbspro.2014.03.224>

Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *The Qualitative Report*, 17(42), 1-10.

Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of assessment for learning in Sweden. *Assessment in Education: Principles, Policy & Practice*, 22(1), 104-121.

<https://doi.org/10.1080/0969594X.2014.970612>

Juchniewicz, J., Kelly, S. N., & Acklin, A. I. (2014). Rehearsal characteristics of “superior” band directors. *Update: Applications of Research in Music Education*, 32(2), 35-43. <https://doi.org/10.1177/8755123314521221>

Koosimile, A. T. (2002). Access negotiation and curriculum change: Lessons from Botswana. *International Journal of Qualitative Studies in Education*, 15(2), 205-223. <https://doi.org/10.1080/09518390110111929>

Kos Jr., R. (2018). Music education and the well-rounded education provision of the Every Student Succeeds Act: A critical policy analysis. *Arts Education Policy Review*, 119(4), 204-216. <https://doi.org/10.1080/10632913.2017.1327383>

Kyaruzi, F., Strijbos, J.-W., Ufer, S., & Brown, G. T. L. (2018). Teacher AfL perceptions and feedback practices in mathematics education among secondary schools in Tanzania. *Studies in Educational Evaluation*, 59, 1-9.

<https://doi.org/10.1016/j.stueduc.2018.01.004>

Lam, R. (2016). Assessment as learning: examining a cycle of teaching, learning, and

assessment of writing in the portfolio-based classroom. *Studies in Higher Education*, 41(11), 1900-1917. <https://doi.org/10.1080/03075079.2014.999317>

Le A. & Guo, H. (2021) The Effect of Results-Based Intergovernmental Transfers in the Educational Accountability System: An Examination of the Race to the Top Program. *Public Performance & Management Review*, 44(4), 785-816.

<https://doi.org/10.1080/15309576.2021.1881802>

Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2013). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment.

Instructional Science, 42(4), 539-559. <https://doi.org/10.1007/s11251-013-9299-9>

Livingston, K., & Hutchinson, C. (2017). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice*, 24(2), 290-307.

<https://doi.org/10.1080/0969594X.2016.1223016>

Malone, E., & Wright, D. (2017). "To promote that demand": Toward a history of the marketing white paper as a genre. *Journal of Business and Technical*

Communication, 32(1), 113-147. <https://doi.org/10.1177/1050651917729861>

Martin, L. (2020). Using Synchronous Formative Feedback to Facilitate Student Growth. *Music Educators Journal*, 107(2), 51-57.

<https://doi.org/10.1177/0027432120976816>

Merriam, S. B. (2009). *Qualitative case study research* (2nd ed.). San Francisco, CA: Jossey-Bass.

McCallum, S. & Milner, M. (2021). The effectiveness of formative assessment: Student

views and staff reflections. *Assessment & Evaluation in Higher Education*, 46(1), 1-16. <https://doi.org/10.1080/02602938.2020.1754761>

Meier, M. E. (2018). Assess What Matters Most: Recommendations for Gathering Information about Student Learning. Section I: Assessment in Art Education-- Building Knowledge. Assessment White Papers for Art Education. *National Art Education Association*.

Miksza, P. (2007). Effective Practice. *Journal of Research in Music Education*, 55(4), 359-375. <https://doi.org/10.1177/0022429408317513>

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd edition ed.). Thousand Oaks, CA: Sage Publication.

Myers, M. J. (2021). Standards-based assessment for secondary choral ensembles: A framework to document student learning. *Arts Education Policy Review*, 1-12. <https://doi.org/10.1080/10632913.2021.1877229>

NAMM Foundation. (2018). Research briefs: Did you know? Retrieved from <https://www.nammfoundation.org/why-music-matters/research-briefs-did-you-know>

National Association for Music Education. (2014). 20 Important Benefits of Music In Our Schools. <https://nafme.org/20-important-benefits-of-music-in-our-schools/>

National Coalition for Core Arts Standards (2014). National Coalition for Core Arts Standards. In *A Conceptual Framework for Arts Learning*.

Nevada Department of Education (2017). Nevada Educator Performance Framework (NEPF).

[http://www.doe.nv.gov/Educator_Development_and_Support/Nevada_Educator_Performance_Framework\(NEPF\)/](http://www.doe.nv.gov/Educator_Development_and_Support/Nevada_Educator_Performance_Framework(NEPF)/)

Ní Bhroin, M. (2021). Assessment of the Visual Arts: Pupils' Voices. Section V: Visual Arts Assessments--Case Studies from the Classroom and Beyond. Assessment White Papers for Art Education. *National Art Education Association*.

Orzolek, D. C. (2020). Effective and Engaged Followership: Assessing Student Participation in Ensembles. *Music Educators Journal*, 106(3), 47–53.

<https://doi.org/10.1177/0027432119892057>

Panadero, E., & Brown, G. T. (2017). Teachers' reasons for using peer assessment: positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133-156. <https://doi.org/10.1007/s10212-015-0282-5>

Partnership for 21st Century Learning (2019a). Assessment of 21st Century Skills.

<http://www.p21.org/storage/documents/Assessment092806.pdf>

Partnership for 21st Century Learning (2019b). Framework for 21st Century Learning.

<http://www.p21.org/about-us/p21-framework>

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage Publications.

Payne, P. D., Burrack, F., Parkes, K. A., & Wesolowski, B. (2019). An Emerging Process of Assessment in Music Education. *Music Educators Journal*, 105(3), 36–44.

<https://doi.org/10.1177/0027432118818880>

Pellegrino, K. (2014). Music-making as data. In C. M. Conway (Ed.), *The Oxford handbook of qualitative research in American music education* (pp. 307-324).

New York: Oxford University Press.

Pellegrino, K., Conway, C. M., & Russell, J. A. (2015). Assessment in performance-based secondary music classes. *Music Educators Journal*, *102*(1), 48-55.

<https://doi.org/10.1177/0027432115590183>

Perrine, W. M. (2013). Music teacher assessment and Race to the Top: An initiative in Florida. *Music Educators Journal*, *100*(1), 39-44.

<https://doi.org/10.1177/0027432113490738>

Perrine, W. M. (2016). Effects of selected nonmusical characteristics and band festival participation, scores, and literature difficulty. *Arts Education Policy Review*,

117(1), 19-28. <https://doi.org/10.1080/10632913.2014.984262>

Platt, J. (1992). "Case study" in American methodological thought. *Current*

Sociology, *40*(1), 17-48. <https://doi.org/10.1177/001139292040001004>

Rawlings, J. R. (2013). An interview with Lynn Tuttle about the Core Arts Standards:

Embracing a new paradigm in arts education. *Arts Education Policy Review*,

114(3), 157-161. <https://doi.org/10.1080/10632913.2013.803408>

Richerme, L. K. (2016). Measuring music education. *Journal of Research in Music*

Education, *64*(3), 274-293. <https://doi.org/10.1177/0022429416659250>

Robertson-Kraft, C., & Zhang, R. S. (2018). Keeping Great Teachers: A Case Study on the Impact and Implementation of a Pilot Teacher Evaluation

System. *Educational Policy*, *32*(3), 363-394.

<https://doi.org/10.1177/0895904816637685>

Roulston, K. (2014). Conducting and analyzing individual interviews. In C. M. Conway

- (Ed.), *The Oxford handbook of qualitative research in American music education* (pp. 250-270). New York: Oxford University Press.
- Russell, B. E. (2015). An empirical study of a solo performance assessment model. *International Journal of Music Education, 33*(3), 359-371.
<https://doi.org/10.1177/02557614155581282>
- Russell, J. A., & Austin, J. R. (2010). Assessment practices of secondary music teachers. *Journal of Research in Music Education, 58*(1), 37-54.
<https://doi.org/10.1177/0022429409360062>
- Sabol, F. R. (2018a). Model Cornerstone Assessments (MCAs): A Powerful Tool for Measuring Student Achievement in Visual Arts Education. Section II: Assessments That Promote Vibrant Learning Communities and Advocate for the Visual Arts. Assessment White Papers for Art Education. *National Art Education Association*.
- Sabol, F. R. (2018b). Some Guiding Principles for Conducting Assessments in Visual Arts Education. Section III: Planning and Implementing Visual Arts Assessments. Assessment White Papers for Art Education. *National Art Education Association*.
- Saldaña, J. (2013). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage Publications.
- Shapiro, E. (1973). Educational evaluation: Rethinking the criteria of competence. *The School Review, 81*(4), 523-549. <https://doi.org/10.1086/443101>
- Sharma, P. (2015). Standards-based assessments in the classroom. *Contemporary Education Dialogue, 12*(1), 6-30. <https://doi.org/10.1177/0973184914556864>

- Shaw, R. D. (2014). An interview with Marcia McCaffrey about the Core Arts Standards: Implications for arts teacher evaluation. *Arts Education Policy Review, 115*(3), 104-108. <https://doi.org/10.1080/10632913.2014.914405>
- Shelden, D. L., Angell, M. E., Stoner, J. B., & Roseland, B. D. (2010). School principals' influence on trust: Perspectives of mothers of children with disabilities. *The Journal of Educational Research, 103*(3), 159-170. <https://doi.org/10.1080/00220670903382921>
- Sickler-Voigt, D. C. (2020a). Assessing Children's Artistic Development through Creative Response Journaling. Assessment White Papers for Art Education. Section V: Visual Arts Assessments--Case Studies from the Classroom and Beyond. Assessment White Papers for Art Education. *National Art Education Association*.
- Sickler-Voigt, D. C. (2018b). Changing Mindsets about edTPA: From Test Anxiety to Demonstrating Teacher Competencies through Authentic Teaching and Assessment Practices. Section II: Assessments That Promote Vibrant Learning Communities and Advocate for the Visual Arts. Assessment White Papers for Art Education. *National Art Education Association*.
- Sickler-Voigt, D. C. (2018c). Introduction. *National Art Education Association*.
- Siebert, J. (2014). NAFME's Teacher Evaluation Workbook. <https://nafme.org/teacherevaluationworkbook/>
- Silveira, J. M. (2013). Idea bank: Portfolios and assessment in music classes. *Music Educators Journal, 99*(3), 15-24. <https://doi.org/10.1177/0027432112470071>

- Smit, R., & Birri, T. (2014). Assuring the quality of standards-oriented classroom assessment with rubrics for complex competencies. *Studies in Educational Evaluation, 43*, 5-13. <https://doi.org/10.1016/j.stueduc.2014.02.002>
- Springer, D. G., & Bradley, K. D. (2017). Investigating adjudicator bias in concert band evaluations: An application of the Many-Facets Rasch Model. *Musicae Scientiae, 22*(3), 377-393. <https://doi.org/10.1177/1029864917697782>
- St. Pierre, N. A., & Wuttke, B. C. (2017). Standards-based grading practices among practicing music educators. *Update: Applications of Research in Music Education, 35*(2), 30-37. <https://doi.org/10.1177/8755123315604468>
- Stevens, D. D., & Cooper, J. E. (2009). *Journal keeping: How to use reflective writing for effective learning, teaching, professional insight, and positive change*: Stylus Publishing, LLC.
- Thompson, M. D., & Penney, D. (2015). Assessment literacy in primary physical education. *European Physical Education Review, 21*(4), 485-503. <https://doi.org/10.1177/1356336X15584087>
- Tomhave, R. D. (2019). The Untold Power of Internationale Baccalaureate Portfolio Assessment in the Visual Arts. Section IV: Analyzing, Interpreting, and Reporting Art Education Assessments. Assessment White Papers for Art Education. *National Art Education Association*.
- Trotman, D. (2006). Interpreting imaginative lifeworlds: Phenomenological approaches in imagination and the evaluation of educational practice. *Qualitative Research, 6*(2), 245-265. <https://doi.org/10.1177/1468794106062712>

- Valle, C., Andrade, H., Palma, M., & Hefferen, J. (2016). Applications of peer assessment and self-assessment in music. *Music Educators Journal*, 102(4), 41-49. <https://doi.org/10.1177/0027432116644652>
- Vidwans, A., Gururani, S., Wu, C.-W., Subramanian, V., Swaminathan, R. V., & Lerch, A. (2017). *Objective descriptors for the assessment of student music performances*. Paper presented at the Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio.
- Wainwright, S. P., Williams, C., & Turner, B. S. (2006). Varieties of habitus and the embodiment of ballet. *Qualitative research*, 6(4), 535-558. <https://doi.org/10.1177%2F1468794106068023>
- Wesolowski, B. C. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, 98(3), 36-42. <https://doi.org/10.1177/0027432111432524>
- Wesolowski, B. C. (2013). Cognition and the assessment of interaction episodes in jazz improvisation. *Psychomusicology: Music, Mind, and Brain*, 23(4), 236-242. <https://doi.org/10.1037/pmu0000016>
- Wesolowski, B. C. (2014). Documenting student learning in music performance. *Music Educators Journal*, 101(1), 77-85. <https://doi.org/10.1177/0027432114540475>
- Wesolowski, B. C. (2015a). Assessing jazz big band performance: The development, validation, and application of a facet-factorial rating scale. *Psychology of Music*, 44(3), 324-339. <https://doi.org/10.1177/0305735614567700>
- Wesolowski, B. C. (2015b). Tracking student achievement in music performance *Music*

Educators Journal, 102(1), 39-47. <https://doi.org/10.1177/0027432115589352>

Wesolowski, B. C. (2016). Exploring rater cognition: A typology of raters in the context of music performance assessment. *Psychology of Music*, 45(3), 375-399.

<https://doi.org/10.1177/0305735616665004>

Wesolowski, B. C. (2017). A facet-factorial approach towards the development and validation of a jazz rhythm section performance rating scale. *International Journal of Music Education*, 35(1), 17-30.

<https://doi.org/10.1177/0255761415590524>

Wesolowski, B. C. (2021). An Examination of Differential Item Functioning in a Rubric to Assess Solo Music Performance. *Musicae Scientiae*, 25(2), 161-175.

<https://doi.org/10.1177/1029864919859928>

Wesolowski, B. C., Athanas, M. I., Burton, J. S., Edwards, A. S., Edwards, K. E., Goins, Q. R., ... & Thompson, J. E. (2018). Judgmental standard setting: The development of objective content and performance standards for secondary-level solo instrumental music assessment. *Journal of Research in Music Education*, 66(2), 224-245.

<https://doi.org/10.1177/0022429418765482>

Wesolowski, B. C., Amend, R. M., Barnstead, T. S., Edwards, A. S., Everhart, M., Goins, Q. R., . . . Williams, J. D. (2017). The development of a secondary-level solo wind instrument performance rubric using the Multifaceted Rasch Partial Credit Measurement Model. *Journal of Research in Music Education*, 65(1), 95-119.

<https://doi.org/10.1177/0022429417694873>

Wesolowski, B. C., & Wind, S. A. (2017). Investigating rater accuracy in the context of

secondary-level solo instrumental music performance. *Musicae Scientiae*, 23(2), 157-176. <https://doi.org/10.1177/1029864917713805>

Wesolowski, B. C., Wind, S. A., & Engelhard, G. (2015). Rater fairness in music performance assessment: Evaluating model-data fit and differential rater functioning. *Musicae Scientiae*, 19(2), 147-170. <https://doi.org/10.1177/1029864915589014>

Wesolowski, B. C., Wind, S. A., & Engelhard, G. (2016). Examining rater precision in music performance assessment. *Music Perception: An Interdisciplinary Journal*, 33(5), 662-678. <https://doi.org/10.1525/mp.2016.33.5.662>

Willerson, A. (2019). Bridging the gap: How the National Core Arts Standards forge the divide between the arts and assessment-based learning. *Arts Education Policy Review*, 120(4), 221-227. <https://doi.org/10.1080/10632913.2018.1443861>

Williams, J. C. (2015). "Assessing Without Levels": Preliminary research on assessment literacy in one primary school. *Educational Studies*, 41(3), 341-346. <https://doi.org/10.1080/03055698.2015.1007926>

Wind, S. A., Engelhard, G., & Wesolowski, B. (2016). Exploring the effects of rater linking designs and rater fit on achievement estimates within the context of music performance assessments. *Educational Assessment*, 21(4), 278-299. <https://doi.org/10.1080/10627197.2016.1236676>

Wong, M. W. Y. (2013). Assessment for learning, a decade on: Self-reported assessment practices of secondary school music teachers in Hong Kong. *International Journal of Music Education*, 32(1), 70-83.

<https://doi.org/10.1177/0255761413491056>

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, 22(1), 140-160. <https://doi.org/10.1080/0969594X.2014.990416>

Xavier University Writing Lab. (2014). How to Write a Position Paper.

https://www.xavier.edu/library/students/documents/position_paper.pdf

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.

<https://doi.org/10.1016/j.tate.2016.05.010>

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L. & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), 228-260. <https://doi.org/10.1080/0969594X.2021.1884042>

Yin, R. (2009). *Case study research design and methods* (4th ed.). Thousand Oaks, CA: Sage Publications.

Zaleski, D. J. (2014). An introduction to classroom assessment for today's music educator. *Illinois Music Educator*, 75(1), 58-59.

Appendix A: The Project

Assessment Literacy of Middle School Music Educators for Music Educators and School

Leadership: A White Paper

by

William Daines

Introduction

Starting with the Nation at Risk report in 1983 and moving through the next two decades of educational initiatives like the Goals 2000 initiative in 1994, the National Assessment of Educational Progress (NAEP) assessments in 1997, and the 2010 Common Core Standards for English Language Arts and Mathematics, the arts disciplines have sought to be a core subject in students' curriculum (Helton, 2020; Willerson, 2019). In the past, there has been a separation between the arts disciplines and assessment-based practices due to arts assessments not being connected with clearly defined standards (Willerson, 2019). The creation of the 1994 arts standards was a first step in providing standards in the arts and in 2014 the standards for arts education were updated to keep pace with the common core standards (Rawlings, 2013).

Along with the 2014 National Core Arts Standards (NCAS) came Model Cornerstone Assessments (MCA) to give educators examples of student work aligned with the new standards (Shaw, 2014). The standards and assessments were created to ensure the assessments of students remain valid to that artistic process (Rawlings, 2014). The NCAS were to be an answer to the problem of arts assessment not being

Recommendations centered on the practices of middle school music educators to assist with (1) Assessment purpose and type alignment, (2) Nonmusical Assessment, and (3) Using Rubrics

clearly connected to standards. In the NCAS, assessment language is used throughout the NCAS documents including verbiage that would be easily identifiable to non-arts policy makers such as measurable, outcome based, rigorous, and sequential approach (Willerson, 2019). The NCAS were developed to help arts teachers use assessment in their classrooms, to further validate the arts as a core subject, and to show the arts will prepare students with 21st century skills (Willerson, 2019).

The increased interest in assessment practices of arts teachers has mirrored the increased interest in the education policy makers' community. Measuring student learning and growth has led the education policy since the start of the Nation at Risk. Along with the increased interest of student learning and growth during this time, there has followed an increased desire to hold teachers accountable for student learning. For this reason, assessment and teachers' assessment knowledge and skill have been the center of the last national educational initiative, Race To the Top in 2009, and included not only how students are achieving and growing but then using that data to evaluate

teachers (Aguilar, 2014).

In this context of heightened awareness of assessments and teacher evaluations, there is a need to support teachers in the development of their assessment knowledge and assessment practices. First, we must know what the current assessment knowledge level of teachers are and their assessment practices. This white paper is the result of a qualitative case study conducted with 12 middle school music educators to identify their current assessment knowledge and assessment practices. Based on the findings of my study, three recommendations are suggested to help music teachers further develop their assessment knowledge and practices. This white paper is for both the middle school music educator and their school leaders. In this paper, I outline the current problem, define assessment literacy, and provide three recommendations for music teachers and school leaders to implement, followed by the educational benefits of an assessment literate teacher.

Problem Definition

In order to examine the current assessment practices of music educators, 12 music educators were interviewed, and their responses were analyzed for commonalities. The findings from the study include three areas where music educators reported they could gain further knowledge about assessments and improve their assessment practices,

(a) grading practices aligned with assessment purposes, (b) grading of nonmusical achievement, and (c) using rubrics more effectively.

In addition to these three areas for improvement, music educators also expressed a need for support from their schools' leaders. Often school leaders do not have a music education background and district or state teacher evaluations do not match what music teachers actually do in the classroom (Robinson, 2017). The challenge for school leaders is to support their music teachers during classroom observations and formal evaluations. They also should know the range of assessment practices that are observable in a typical music teaching environment. Supervisors should provide meaningful feedback to help all teachers better their craft. If the supervisor demonstrates a gap in understanding the discipline, then teachers might feel the need to "fake teach" or engage in "hoop jumping" or inauthentic teaching (Robinson, 2017) when observed to guarantee an evaluation score of four out of four. This type of experience decreases the opportunity for growth on the part of the teacher, including the development of assessment knowledge and practices.

This white paper was drafted to address the needs of music teachers and their school leaders. The recommendations from the study include ways for music teachers to align their assessment practices with the current

literature and gives school leaders a “what to look for” when observing a music teacher. This approach of both teacher-initiated internal evaluation and external observations by a supervisor of assessment practices will allow music teachers to develop their assessment literacy by making these practices a part of their teaching and not another education initiative that loses momentum over the long term.

Assessment Literacy

Assessment literacy is “the skills and knowledge teachers require to measure and support student learning through assessment” (DeLuca, 2016, p. 248) as well as how these fit into all aspects of professional education culture, teacher identity, and classroom practices (Coombs, 2018). There are many pieces which fit together to create an educator’s assessor identity including the knowledge base, interpretive and guiding framework, conception of assessment, macro and mirco contexts, assessment practice, teacher learning, teacher reflection, teacher participation, and reconstruction of assessor identity (Xu & Brown, 2016). The purpose of the study was to better understand middle school music educators’ assessment knowledge and assessment practices.

Assessment Knowledge

Assessment literacy knowledge is necessary as it is used to provide a

standard for good assessment (Xu & Brown, 2016). The knowledge base in assessment is the disciplinary knowledge and pedagogical content knowledge of the teacher. The curriculum and standards of the discipline are the guides upon which assessments are measured. Teachers must understand the purpose, content, and methods of assessment, how to connect assessments to learning goals, and why assessments are necessary. Assessment knowledge includes an understanding of grading rationales for grading, and various ways of grading assessments with the use of rubrics and criteria. Another component is the knowledge of feedback which includes the principles of feedback and

“A teacher's approach to assessment is comprised of both conceptual understandings and practical knowledge related to student assessment within the situated context of their

the various ways of providing feedback. Once assessments are administered, teachers need to know how to interpret the data and how to communicate results to stakeholders. Teachers should know how to use self and peer assessment with students and how to include students to participate in assessment. The final component is the knowledge of assessment ethics, which is teachers’ responsibilities in sharing and storing assessment data, how to develop and sustain equity, social justice,

nondiscrimination, and inclusion in the creation and administration of assessments. The development of a teacher's assessment literacy includes an understanding and knowledge of these areas to correctly develop the assessment practices of the teacher (Xu & Brown, 2016).

Assessment Practices

The assessment practices of teachers involve the decision making and action taking processes teachers make in the classroom. This process is the teachers' ability to manage external factors—like class size, time, teaching schedule, etc.—using the teachers' knowledge, beliefs, and conceptions about assessment. For example, a mathematics teacher can easily have a “ticket out the door” assessment at the end of class to find out if new content was learned during class. Band teachers who have over 60 students in one class and teaching up to 10 different instruments during that same period would have a more challenging task of creating an exit ticket which coincides with students and their instruments, distributing the exit ticket, and then grading the exit ticket in a timely manner. A middle school music educator would use their knowledge about assessment to balance the assessment of student learning with the context the assessments are taking place. Music teachers' assessment practices must take into consideration the context which

assessments are administered along with the grading, feedback, and communication of the data.

Assessment Recommendations

The following recommendations focus on three areas of assessment knowledge and practices and are based on the findings from the study. Middle school music educators can use these three recommendations as a self-check to further develop their assessment literacy. Teachers can discuss with a colleague or reflecting on their own assessment knowledge and practices. It is important to take into consideration that change, and development occurs over time and the challenge of change must be met with wisdom and balance. Change takes patience to create new ways of viewing assessment and these moments of reflection will help develop a new more extensive assessor identity.

School leaders can use these recommendations to support music teachers in extending their assessment practices. School leaders might also use these during the evaluation process when discussing goals with music teachers and what goals would be meaningful to music teachers. School leaders can support music teachers by making it possible for a group of music teachers to meet and discuss assessment literacy, sharing assessment ideas, practices, positive and negative experiences, and seek input from colleagues. School leaders have influence to direct the

growth of their music teachers and to provide a way for them to develop their assessment literacy.

Recommendation #1: Alignment of Grading Practices with the Assessment Purpose

The first recommendation is the alignment of grading practices with the assessment purpose. The purpose of the assessment determines whether the assessment is formative or summative. For example, if the assessment is to find how students are doing during the unit of learning, the assessment purpose would align with formative assessment. If the assessment is to see what students have learned at the end of a unit, that assessment purpose would align with a summative assessment. It is important that the purpose and type of assessment are aligned to give the correct weight to each type of assessment in the grade book, for the data to be interpreted correctly by the teacher, and for the correct type of feedback being given to students. Music teachers need to ensure that formative assessment has its place of guiding instruction (Livingston & Hutchinson, 2017) and to place its weighted value in the grade book as lesser than when compared with summative assessments.

Some music educators in this study wanted to place an importance on practicing, a formative assessment, and incorrectly weighted these the same as a playing assessment, which would be

summative. This participant shared the categories of his grade book as “four major categories: practice logs, participation, playing tests, and a [weekly] worksheet,” which placed practice logs the same as playing assessments. While practicing and creating a practice habit for students is extremely important, the activity is preparing the students for the playing assessment which is a summative assessment. The practicing assessment should be labeled as a formative assessment and weighted less than summative assessments in the teacher’s grade book. Teachers who wish to place an importance on practicing could consider other methods, like portfolios, that place the importance on practicing they desire while correctly aligning the purpose with the assessment type.

An example of how to align practice assessments with the correct purpose would be to use portfolios. Wesolowski (2014) documented how music educators can use portfolios in the diagnostic assessment, formative assessment, and summative assessment of students’ learning. The strength of the portfolio is the opportunity it gives for teachers to individualize or differentiate learning for students. They also provide music educators the ability to collect various forms of data to improve planning and instruction (Wesolowski, 2014). Over time students could create a portfolio of recordings each week for self-reflection and assess how they are

progressing toward their goals. At the end of the specific time period, students complete a final recording and submit their portfolio of recordings. By using portfolios, a music teacher can guide instruction with formative assessments and identify what learning has occurred with a summative assessment. In this way, the importance of practicing is emphasized, and purposes of assessments are aligned correctly with the assessment type.

Recommendation #2: Grading of Nonmusical Achievement

The second recommendation is what decision is best for the grading of nonmusical achievement like students' attitude, attendance, and preparedness (students have materials in class like music, instrument, pencil, etc.) often labeled as participation points. The grade a student earns during the grading period should reflect the student's musical achievement (Sharma, 2015). However, some music educators use participation points (Pellegrino, Conway, & Russell, 2015) in the gradebook like a musical achievement grade which inflates the real achievement of the student. From this study, one music educator placed a high importance of participation points by making it equal to performances, "a concert grade is weighed at 40% of their grade and so is participation." Another

mentioned, her grade book weight as "40% is daily participation and practice logs."

For accuracy, reliability, and validity, music teachers need to use a standards-based grading approach for student assessments (St. Pierre & Wuttke, 2017). Music teachers need to place nonmusical assessments as a separate citizenship grade since they are not musical standards. By doing this, the teacher is communicating to the student and parents at what level the student is achieving musical skills (letter grade)

Rubric "criteria that provide qualitative descriptions that authentically and accurately distinguish performance levels are valuable. These criteria enable feedback that will support learning and improve performance"
DeLuca & Bolden (2014)

and at what level is the student being responsible in the classroom (citizenship grade). This provides clarity on at what level musical skills are being achieved by the student and at what level the students is demonstrating nonmusical skills.

Recommendation #3: Use of Rubrics in Grading Assessments

The final recommendation concerns how music educators use rubrics in grading assessments. Rubrics are an effective way to assess the

complexities of a musical performance (DeLuca & Bolden, 2014) and provide music educators a quick method for scoring many students with high quality (Doğan, 2017). The criteria selected is critical for the effectiveness of the rubric. Rubric construction needs to be founded in standards which are clear and useful in assessment and feedback (DeLuca & Bolden, 2014; Smit, 2014). The use of learning goals and purpose of the assessment will also ground the rubric with the clarity and specificity needed (Wesolowski, 2012).

Music educators use their expertise to select the criteria for each assessment. An important aspect to selecting criteria is understanding that choosing a certain set of criteria for one assessment will exclude or minimize other musical criteria (Richerme, 2016). Therefore, the criteria which music teachers select to communicate to students what the most important musical skill students should focus on in their personal practice and diminishes the importance of the criteria not selected. Also, if a rubric has all components of music skills included and is the only rubric ever used, it fails to place any criteria as important and does not communicate to students what skills to practice.

Music educators must understand the needs of the students and select criteria which focuses on those needs. The participants in the study used rubrics for grading but had a rubric they used

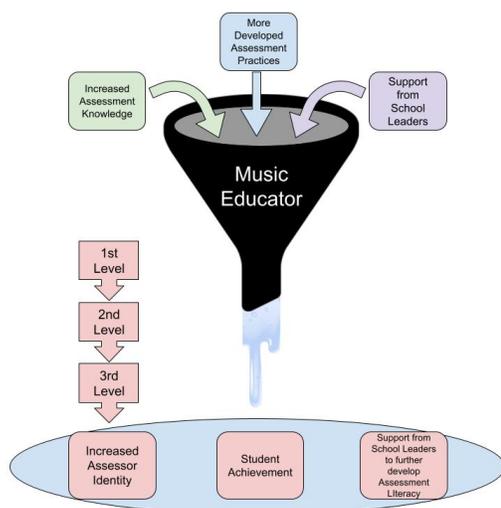
throughout the year without changing the criteria. Many followed what this participant said, "I use the festival rubric we have for choir. They have five categories and we go through those categories [as a class]." Using the same rubric for every assessment reduces the effectiveness of the rubric. By using the same rubric for every playing assessment, students do not receive a clear message from the teacher what is the most important skill or concept to be focusing on in their personal practice and the growth of the student may be hindered. There is a more effective method for using rubrics to help students get a clear message of what is most important to promote growth.

An example of using rubrics in a more effective way would be if a band teacher noticed a group of students were not articulating notes correctly. That teacher could place on the rubric for the next assessment criteria such as, correct tone, correct start to each note, and correct end to each note. Placing this focus on each element of articulation allows for the teacher to give meaningful feedback and for students to know what important musical concept needs work. This would be communicated to the students prior to the assessment and gives the students time to focus on this concept during their preparation. During the next assessment, the band teacher notices that executing 16th notes needs improvement and will change the rubric criteria to match that musical concept.

This method focuses on the needs of the students, shows the student how the passage needs to be performed, and communicates what is important.

Educational Benefits of the Recommendations

Using the recommendations from the study directly benefit the main stakeholders in education, namely students, teachers, school leaders, and less directly the parents. Music educators who integrate these recommendations into their practices will further develop their assessment literacy which benefits students. Students will benefit from a teacher who is assessment literate by having higher achievement and growth. School leaders will increase their knowledge of what assessment practices are in a music classroom and can then give meaningful feedback to music educators during classroom observations.



Students will benefit from music teachers using the recommendations to develop their assessment literacy. As teachers reconstruct their assessor identity through assessment literacy development, they progress through three different levels of assessment mastery (Xu & Brown, 2016). Teachers move from mastery of the fundamental principles of assessment to the teachers' perceptions of assessments shaping how assessments should be. The final stage is a self-awareness of teachers which allows them to self-reflect and gain new insights in their own assessment practices (Xu & Brown, 2016). This growth process is important. As teachers gain a more developed assessment literacy mastery, it enables the teachers to have a greater impact on student learning, growth, and achievement.

Teachers will benefit from school leaders who use the recommendations to better understand the assessment practices of music educators. School leaders can give more meaningful feedback during observing and giving evaluations of music educators. Music teachers have felt that evaluations lack meaning since the school leader was not familiar with the expectations of a music classroom and this “created tensions that manifested in diminished feelings of self-efficacy, classroom control, and decreases in personal and professional worth and value” (Robinson, 2017, p. 52). Once school leaders are more versed in what to look for as they are

observing in a music classroom and can give meaningful feedback, teachers can then receive the support needed to more fully develop their assessment literacy.

Summary

The recommendations from this paper are based on the research findings of a qualitative study. The recommendations of (a) grading practices aligned with assessment purposes, (b) grading of nonmusical achievement, and (c) using rubrics more effectively are to help music teachers more fully develop their assessment literacy. School leaders who know about these recommendations and understand how assessment literacy is developed can support music teachers in their goals to become more knowledgeable about assessment and effective assessment practices. The benefits of following these recommendations include an increased assessor identity, greater student achievement, and school leaders who can support the development of their music educators.

References

- Aguilar, C. E., & Richerme, L. K. (2014). A descriptive study of music teacher educators' beliefs about policy. *Journal of Music Teacher Education*, 25(2), 37-49. <https://doi.org/10.1177/1057083714553986>
- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134-144. <https://doi.org/10.1016/j.tate.2017.12.010>
- DeLuca, C., & Bolden, B. (2014). Music performance assessment. *Music Educators Journal*, 101(1), 70-76. <https://doi.org/10.1177/0027432114540336>
- DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016). Teachers' approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, 25(4), 355-375. <https://doi.org/10.1080/0969594X.2016.1244514>
- Doğan, C. D., & Uluman, M. (2017). A comparison of rubrics and graded category rating scales with various methods regarding raters' reliability. *Educational Sciences: Theory & Practice*, 17(2), 631-651. <https://doi.org/10.12738/estp.2017.2.0321>
- Helton, B. (2021) The arts' legitimacy problem. *Arts Education Policy Review*, 122(4), 224-238. <https://doi.org/10.1080/10632913.2020.1731898>
- Livingston, K., & Hutchinson, C. (2017). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice*, 24(2), 290-307.

<https://doi.org/10.1080/0969594X.2016.1223016>

- Pellegrino, K., Conway, C., & Russell, J. (2015). Assessment in performance-based secondary music classes. *Music Educators Journal*, 102(1), 48-55.
<https://doi.org/10.1177/0027432115590183>
- Rawlings, J. R. (2013). An interview with Lynn Tuttle about the Core Arts Standards: Embracing a new paradigm in arts education. *Arts Education Policy Review*, 114(3), 157-161. <https://doi.org/10.1080/10632913.2013.803408>
- Richerme, L. (2016). Measuring music education. *Journal of Research in Music Education*, 64(3), 274-293. <https://doi.org/10.1177/0022429416659250>
- Robinson, M. (2019) Music teachers' perceptions of high stakes teacher evaluation. *Arts Education Policy Review*, 120(1), 45-56.
<https://doi.org/10.1080/10632913.2017.1373380>
- Sharma, P. (2015). Standards-based assessments in the classroom. *Contemporary Education Dialogue*, 12(1), 6-30. <https://doi.org/10.1177/0973184914556864>
- Shaw, R. D. (2014). An interview with Marcia McCaffrey about the Core Arts Standards: Implications for arts teacher evaluation. *Arts Education Policy Review*, 115(3), 104-108. <https://doi.org/10.1080/10632913.2014.914405>
- Smit, R., & Birri, T. (2014). Assuring the quality of standards-oriented classroom assessment with rubrics for complex competencies. *Studies in Educational Evaluation*, 43, 5-13. <https://doi.org/10.1016/j.stueduc.2014.02.002>
- St. Pierre, N. & Wuttke, B. (2017). Standards-based grading practices among practicing music educators. *Update: Applications of Research in Music Education*, 35(2),

30-37. <https://doi.org/10.1177/8755123315604468>

Wesolowski, B. C. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, 98(3), 36-42.

<https://doi.org/10.1177/0027432111432524>

Wesolowski, B. C. (2014). Documenting student learning in music performance. *Music Educators Journal*, 101(1), 77-85. <https://doi.org/10.1177/0027432114540475>

Willerson, A. (2019) Bridging the gap: How the National Core Arts Standards forge the divide between the arts and assessment-based learning, *Arts Education Policy Review*, 120(4), 221-227. <https://doi.org/10.1080/10632913.2018.1443861>

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.

<https://doi.org/10.1016/j.tate.2016.05.010>

Professional Development Evaluation Survey

1. Name of professional learning session:
2. Date of professional learning session:
3. Name of Presenter:
4. Your name:
5. Level:
6. Employee category:
7. Name of school:
8. Select the division that presented this session:
9. The knowledge learning in this professional learning will support me in effectively implements digital learning practices: (Agree/Disagree)
10. The knowledge learning in this professional learning will help me meet the needs of students. (All students/English language learners/Special education students)
11. The knowledge learned from this professional learning will be utilized in planning and instruction. (Agree/Disagree)
12. Do you plan to implement a strategy that was shared, developed, learned, or planned during this session? (Yes/No)
13. What strategy covered during the session will you include in your practice within a week?
14. Please add any additional comments or questions below:
15. What additional support do you need?
16. What follow-up professional learning would you like to suggest?

Appendix B: Permission to Use the Teacher Assessment Literacy in Practice Illustration

**ELSEVIER LICENSE
TERMS AND CONDITIONS**

Jun 27, 2018

This Agreement between William Daines ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4377151201778
License date	Jun 27, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Teaching and Teacher Education
Licensed Content Title	Teacher assessment literacy in practice: A reconceptualization
Licensed Content Author	Yueting Xu,Gavin T.L. Brown
Licensed Content Date	Aug 1, 2016
Licensed Content Volume	58
Licensed Content Issue	n/a
Licensed Content Pages	14
Start Page	149
End Page	162
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Figure 1
Title of your thesis/dissertation	Assessment with Middle School Music Educators
Expected completion date	Dec 2018
Estimated size (number of pages)	60
Requestor Location	William Daines

Appendix C: Interview Protocol

“Hello! My name is Will Daines and I am an Ed D student at Walden University in the curriculum, instruction, and assessment program. Thank you for your participation in this study and your willingness to be interviewed today. This interview is part of a project study for me to complete my Ed D degree. For the project study I am focusing on the assessment knowledge and practices of middle school music educators. The purpose of this interview is to learn about assessment knowledge and practices among middle school music educators. There are no right or wrong answers, or desirable or undesirable answers.

“If it’s okay, I will be recording the audio of our interview with my tablet and later transcribe this interview myself. The transcript will be used only for the purposes of this study and will remain private. At any point in time, you can choose to stop the interview or choose to not answer a question. Everything you say will be confidential. Myself and my committee chair will be the only ones who are aware of your answers. At the end of the interview I will collect your two assessment samples we had discussed previously. When I have completed the data analysis for this study, I will provide you with the findings for your information. What questions do you have before we begin? (wait for any questions and answer any questions before proceeding) Again, thank you for your participation today and we will get started.”

During the interview I understand the participant might feel apprehensive. I must create an environment which will help the participant to get from apprehension to open participation. Methods to do this would be to be friendly and non-judgmental in my

mannerisms. Examples of follow up questions to avoid are “what do you mean by that?” or “why would you do that?” which have a hidden meaning of being judgmental. As the interviewer, listening will be an important skill since this will lead to better probing questions and, therefore, gathering better data (Merriam, 2009). Throughout the interview, I will use follow up questions which probe for further details help to clarify answers or to learn more about their experience around classroom assessment. At the end of the interview, participants will be thanked for their time and their valuable input. I will make sure the participant has my contact information for any concerns or questions he or she may have.

Interview Questions

1. Describe your professional background.
2. What does assessment mean to you?
3. Throughout your teaching career, how has your understanding of assessment changed?
4. What is the purpose of your music assessment for band, choir, or orchestra?
5. What outside influences effect your choice of assessment?
 - a. Describe a time when you had to change your choice of assessment based on an outside influence?
 - b. How do you decide which assessment type to select for assessing student skills and student knowledge in band, choir, or orchestra?
6. What are the most common assessments you use in your classroom? Why do you use them?

- a. Describe the formative assessments you developed and use to determine students' knowledge and skills in band, choir, or orchestra.
 - b. Describe the summative assessments you developed and use to determine students' knowledge and skills in band, choir, or orchestra.
7. How do you give feedback to students and parents?
 - a. How do you provide formative and summative feedback when assessing students' knowledge and skills in band, choir, or orchestra?
 - b. How do you relay this feedback to students' parents?
8. What do you do with the results of assessments? How do you use assessment results for planning next steps for students' knowledge and skills in band, choir, or orchestra?
9. How do you use assessment results to determine students' strengths and weaknesses in music knowledge and performance skills in band, choir, or orchestra?
10. What are your grading practices?
 - a. How do you determine your grading practices in band, choir, or orchestra?
 - b. Describe how you explain your grading practices to your students.
11. How might a student describe the assessments you use in the classroom? How do you account for bias when grading students?
12. How do you include students in classroom assessments?
 - a. What information have you gleaned from involving students in developing classroom assessments?

- b. How have you modified your instruction or assessment practices based on student involvement in classroom assessment?
- 13. Do you use rubrics or portfolios? If so, are they aligned with the national music standards?
 - a. What steps did you take to ensure the rubrics and portfolios were aligned with the national music standards?
 - b. How do you ensure you are meeting district and state music requirements?

Sample Follow-Up Questions:

1. Tell me more about...
2. You mentioned ..., tell me more...
3. What do you mean by...?

Appendix D: Document Protocol

Document Item #:	Author:
Type of Document:	
Context:	
Purpose:	
Intended Audience:	
Main Points:	
Comments:	

Appendix E: A Priori Codes and Examples

Category and code	Participant	Example
The knowledge base Knowledge	BAND02	We want the mastery of the skill
	BAND03	. . . a proper assessment tool really helps identify the strengths and the deficiencies of a student.
	CHOIR01	Assessment is basically a test on do they understand the knowledge.
Disciplinary knowledge and pedagogical content knowledge	ORCH02	I will work more on intonation in the classroom . . . or I will change what I'm doing in order to help them with the intonation.
	ORCH01	So all violas had to play together. . . Are they all playing the right note at the right time and teamwork?
	CHOIR03	I think that as music teachers, we do it [assess] every day because you have to be listening, and if you aren't listening the kids are going to be singing or playing the wrong notes.
Knowledge of assessment ethics	BAND01	I feel challenged by that because I know we all have biases . . .so I try to grade them based on just the rubric and not any sort of extra stuff.
	CHOIR03	Formal testing I take the bias out . . .you either do it or you don't.
	ORCH02	I just grade what comes out of the instrument, what the fingers do. I just look at the kid for what they're able to do.
Knowledge of assessment interpretation and communication	ORCH03	Assessment is finding out how the students are doing so that I can change what I'm doing or tweak what I'm doing so that they can learn.
	BAND04	. . . [using assessment data to] guide instruction. Most certainly you have to know what they can do and what they can't do and what they need to learn.
	CHOIR02	Assessment, is simply seeing where a student is, seeing what have they learned, how far and where do I need to take them?
Knowledge of assessment purposes, content, and methods	BAND03	Assessment to me, if used properly, should help guide students and help show progress over a period of time.

(table continues)

Category and code	Participant	Example
Knowledge of student involvement in assessment	CHOIR03	I think that as music teachers, we do [assessments] every day. Because you have to be listening, and if you aren't listening the kids are going to be singing or playing the wrong notes. So you're constantly engaged.
	ORCH01	. . . at every kid . . . [Do] they get it, no, they don't, [I] must remediate now.
	CHOIR04	They have to grade themselves before I grade them. And then we compare, we talk about . . . this is what you saw or you thought, this is what I saw.
	CHOIR03	We'll record [the assessment] and then I'll let the kids listen to it. They can give feedback and see if they're hearing the same thing that I've been telling [them] for the last three weeks.
	CHOIR01	I have the different [choir] sections listen to each other sing and I say to them "You're listening for this proper vowel and... for these notes, tone, the blend, the balance.
Teacher assessment literacy in practice Decision-making	ORCH04	I do process work all the time. I break everything down . . . to it's . . . core, . . . and then they start building everything up from just the fundamentals.
Action	CHOIR04	I make sight reading factory an assessment . . . I think that's what shows that the students are actually learning something in your class.
	BAND04	That's what creates the quality . . . every week [the students] perform in front of each other.
	BAND01	If a lot of the kids don't do well on a test, I know that there's something that I missed in teaching it to them. We need to go back and we need to maybe approach it a different way.
	BAND04	If I'm hearing some weird consistency in the clarinet section, I have to pick a warm up that address[es] that. It's more like this isn't being done very well. We need to include this in a warm up.
	CHOIR03	I break it up into chunks and then scaffold them.

Appendix F: Open Codes and Examples

Code	Participant	Example
Academic achievement	BAND01	We're going to try something new this year because of the way that they've asked us to change the grading. We're going to make our winter and spring concerts, their semester exams. It's the first time we've done it.
	BAND03	our classes are broken down into, into four major categories. Practice logs, participation, . . . playing tests. And then once a week . . . we do one worksheet.
	CHOIR02	My concert is the most important thing . . . so that becomes summative and is 300 points. When everything else is under 50 points I put the importance on what I think is important and if that child learns that, then that's where their grade comes from. But that child earns their grade. I write down what the child earns, but I give them this list of this is what is required.
	CHOIR03	All written quizzes are anywhere from five points to 49 points. And then all performing quizzes are 50 points and that's across the board for all three departments Our administrator this year has said that our concerts can be our semester exams. So the winter concert and the spring concert is 20% of their semester grade.
	CHOIR04	We're required to have at least one assessment in our grade books a week at my school. They [the students] get a Monday, Tuesday, Wednesday, Thursday, Friday paper. Each day has something different in it. They fill it out, they answer it, we go over the answers, it's their job to make sure the answers are correct.
	ORCH02	they get a practice log, it's due every Monday. So that's their homework. And then I try to do one assessment per week.
	ORCH03	I do really quick playing quiz. It usually only takes about half the class period and it's just a few measures.

(table continues)

Code	Participant	Example
Participation points	BAND03	Our classes are broken down into, into four major categories. Practice logs, participation, AKA citizenship, things like book check, pencil check, being in class on time. And then they're playing tests.
	BAND04	They show up with their instrument prepared to play and they're actually engaged.
	CHOIR01	Participation, meaning, are they singing in class, are they putting forth effort or are they sitting there doing nothing? And I assess that on an individual, individual basis. A concert grade is weighed at 40% of their grade and so is participation 40%
	CHOIR03	Our participation has demonstration of skills, which is 25 points a week. And that's band, orchestra and choir.
	ORCH02	one of my assignments is just classroom participation. Did you show up, did you do your job, were you prepared, you had your instrument, your rock stop, your shoulder, rest, your music. They get 20 points every day for that.
	ORCH03	And then the 40% is daily participation and practice logs.
	ORCH04	Preparation, if they have pencils on our stands, shoulder rests for the upper strings, and then books all their books, all their music, I'll check it every day.
	Playing or singing assessment	BAND01
BAND02		Because it's so much group play all the time. It's easy for kids to hide. So individual assessment takes that away.

(table continues)

Code	Participant	Example
BAND03		Symphonic band who is in festival prep mode, all of those assessments are playing checks Our classes are mainly performance-based . . . if they can play lights out at the concert or play lights out a festival . . . that's what's most important.
BAND04		I assess more often now . . . they have to play in front of each other They have to play. What they can do on a paper doesn't really necessarily transmit to what we actually do, which is a performance based class. So there has to be playing. That's our purpose.
CHOIR01		What I personally try to do is a formal singing test assessment individually, once per quarter
CHOIR02		Every time I ask a question, I'm assessing, every time I ask you to do something I'm assessing . . . individually, you sing this note - all of that is assessment.
CHOIR03		For my advanced kids it varies. Sometimes there'll be written quizzes, sometimes there'll be quartet quizzes, sometimes there'll be groups singing quizzes or memorization, or ear training
CHOIR04		I make sight reading factory an assessment because I believe in sight reading. I think that's what shows that the students are actually learning something in your class
ORHC01		We had a test last week and beginning orchestra where it was a group test. So all violas had to play together. I'm not measuring anyone's percent specific tone. Are they all playing the right note at the right time and teamwork?
ORCH02		For the intermediate and the advanced, I try to do at least twice a month in individual assessment. Because they need to get used to playing by themselves and hearing the sound that they're making.

(table continues)

Code	Participant	Example
Rubrics	ORCH03	I do very quick assessments every week. I assessed the kids at the end of the week usually. And I do really quick playing quiz. I like the quick little assessments that I can do very often. Playing tests. Lot of playing tests.
	ORCH04	I start hearing students, especially when they're doing playing tests and stuff, I can see, okay, they don't understand this
	BAND01	I modeled my rubric on the festival education sheet. I used the same categories and criteria based on that.
	BAND03	My rubric is a combination of former band directors, current band directors, myself, colleagues back in the Midwest.
	CHOIR02	I use the festival rubric we have for choir. They have five categories and so we go through those categories.
	CHOIR03	For all of my sight-reading quizzes and all my vocal testing, yes, I do rubrics.
	ORCH02	I have a rubric that I use and I write comments on the rubric if they lost a point here or there. I try to write quickly why they lost the point before I move onto the next student. I wrote the rubric myself. Some of it does have [district] standards and some of it does have the national standards.
	ORCH03	I have a playing test rubric that I use . . . most of the time I have a list on the board, I'll just say these are the things I'm looking for.
	ORCH04	I write the rubrics because if my supervisor were to come in, she's going to want to see it . . . But the basic stuff that we learned from the beginning is always going to be on the rubric.
	Weekly assessments	BAND03

(table continues)

Code	Participant	Example
BAND04		I changed to weekly tests a few years ago and just made a tremendous difference But that's what creates the quality is the, every week we perform in front of each other with that assessment
CHOIR01		Their theory that they do every day, it goes into a grade every two weeks. They just keep it in their binder for two weeks and it gets collected every other Fridays that goes in for a grade.
CHOIR03		When I first started teaching, I would give a quiz every two weeks. And now I'm giving one pretty much every week
CHOIR04		We're required to have at least one assessment in our grade books a week at my school
ORCH02		And then I try to do one assessment per week.
ORCH03		I've evolved into is very quick assessments more often. I do very quick assessments every week. I assessed the kids at the end of the week usually
ORCH04		I do weekly tests every week. They [students] have practice logs, their guided practice log . . . two hours in the week

Appendix G: Axial Codes and Examples

Code (Category)	Participant	Example
Alignment in purpose, type of assessment, and grading (Knowledge)	BAND01	The very first test to do is just to play their mouthpiece . . . demonstrated correct embouchure . . . then there's second test is to play the first note that they learned.
	CHOIR03	Our music department is 100% [summative]
	BAND03	Our classes are broken down into four major categories: practice logs, participation, citizenship, . . . playing tests.
Student involvement in assessment (Decision-making and action taking)	CHOIR04	After they watched themselves . . . they talked about how would you improve your performance? What personal goal can you set for your future performance and what do you need to reach that goal?
	ORCH04	I'll ask students, okay, what did you think?
	ORCH02	If I don't feel like the kids have done well enough; [if it's not their best, I'll redo it [the assessment].
	ORCH03	And then I'll ask the kids questions like, what did you hear? What are some things they can improve on?
Assessment routines (Knowledge)	BAND01	I do informal assessments . . . daily with them. I know where we are and if they're ready for that summative assessment or not.
	ORCH03	I do very quick assessments every week.
	BAND03	And every Friday they turn the practice log in. They are required [to practice] two hours a week and they get 10 points towards their weekly grade.
Skills-based feedback (Knowledge)	BAND02	[I give to students] immediately right after [the playing assessment] what was good or what was not, what they need to work on and focus on for the next time.
	ORCH01	Written feedback [is] typically on Google classroom: [this includes determining] at 17 seconds if the wrist was caved in [and needs to be straight].
	ORCH04	And they'll raise their hand, they'll say, F sharp is out of tune.

Appendix H: Themes

Axial category	Theme and subtheme
Alignment in purpose, type of assessment, and grading	Middle school music educators used their knowledge of assessments to assess student performances.
Assessment routines	Aligning assessment purpose to type of assessment.
	Aligning assessments to purpose and grading practices.
Student involvement in assessment	Middle School music educators use knowledge of assessment grading and student involvement feedback
Skills-based feedback	practices to improve student performance.